DOCUMENT RESUME

ED 327 565 TM 015 995

AUTHOR DeStefano, Lizanne; Wagner, Mary
TITLE Outcome Assessment in Special Education: Lessons Learned.
INSTITUTION Illinois Univ., Champaign. Secondary Transition Intervention Effectiveness Inst.; SRI International, Menlo Park, Calif.
PUB DATE 91
NOTE 69p.; Photoreduced print in figure 5 will not reproduce legibly.
PUB TYPE Reports - Evaluative/Feasibility (142) --

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Data Analysis; Data Collection; *Educational Assessment; Elementary Secondary Education; Needs Assessment; *Outcomes of Education; Program Evaluation; *Research Methodology; Sampling; *Special Education
IDENTIFIERS *National Long Transition Study Spec Students

ABSTRACT

The use of outcome assessment in special education is reviewed. Outcome assessments differ widely in design, measures, and data collection, but they share a common focus on outcomes as individual achievements, statuses, or behaviors. Assessments of special education outcomes often have an evaluative purpose in reflecting how well the special education system in general is performing. The assessment process consists of the following sequence of key activities: (1) identifying key issues and information needs; (2) developing a conceptual framework to guide the assessment; (3) specifying the nature of comparisons to be made; (4) designing and selecting a sample; (5) selecting and operationalizing outcome measures; (6) choosing independent variables to illustrate outcome variations; (7) selecting data sources and collection methods; (8) choosing appropriate analysis methods; and (9) communicating findings to encourage their use in policy making and program planning. Each of these activities is described in detail, and examples are drawn from various projects, with emphasis on the National Longitudinal Transition Study of Special Education students, a study of over 8,000 special education students aged 13 to 21 years. Five tables and six figures illustrate the discussion. A 60-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED327565

TRANSITION INSTITUTE AT ILLINOIS

National Transition Study

OUTCOME ASSESSMENT IN SPECIAL EDUCATION: LESSONS LEARNED

Lizanne DeStefano
University of Illinois at Urbana-Champaign

Mary Wagner
SRI International

1991

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MARY WAGNER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

University of Illinois
at Urbana-Champaign

Secondary Transition Intervention Effectiveness Institute

110 Education Building
1310 South Sixth Street
Champaign, Illinois 61820

TM015995

SRI International
333 Ravenswood Ave. • Menlo Park, CA 94025-3493

CONTENTS

What Is Outcome Assessment? 3

The Process of Assessing Outcomes 4

Identifying Key Issues and Information Needs 5

Developing a Conceptual Framework 7

Specifying the Nature of Comparisons To Be Made 11
- Comparisons with the General Population of Youth 11
- Comparisons of Youth with Different Types of Disabilities 13
- Cross-Unit Comparisons 14
- Longitudinal or Time Series Comparisons 16

Designing and Selecting a Sample 16
- What Group(s) Should the Sample Represent? 17
- Sample Size Considerations 18
- Sample Selection Methods 19
- Locating Respondents and Obtaining the Data 22
- Documenting the Generalizability of the Sample 25

Selecting and Operationalizing Outcome Measures 27
- Common Measures of Outcomes for Secondary School Students with Disabilities 28
- Common Postschool Outcome Measures for Youth with Disabilities . . 36

Choosing Independent Variables to Illuminate Outcome Variations 41

Selecting Data Sources and Collection Methods 42
- Alternative Data Sources 42
- Data Collection Methods 47
- Instrument Development 48
- Timing of Data Collection 49

Choosing Analysis Methods 49
- The Nature of Research Questions Asked 50
- The Characteristics of Important Variables 52
- Sample Size and Composition 54
- Knowledge Base and Experiences of Audiences 55

Communicating Outcome Information 56

Outcome Information in Use: Opening Pandora's Box 58

References 60

OUTCOME ASSESSMENT IN SPECIAL EDUCATION: LESSONS LEARNED

The educational reform initiatives that have dominated educational policymaking during the last decade have been accompanied by raised expectations, higher standards, and increased performance accountability for our schools. As financial constraints have tightened, those responsible for providing resources for public education have started to demand more direct evidence of the return on their investment. Legislators, governors, and state and local boards of education have responded to these concerns by focusing on the outcomes associated with education. Such a focus has resulted in a number of outcome-oriented evaluations in education.

As all levels of authority have taken a greater interest in allocating limited education resources to ensure maximum effectiveness, special education students and programs are included in outcome assessments more frequently than in the past. State and federal legislators, practitioners and families have expressed concern about the educational, occupational, and independent living status of individuals with disabilities after leaving school, and the impact of special education programming on those outcomes. These groups also have stated a need to measure the educational skills and outcomes that students attain during their school careers. These interests were recognized in 1983, when the U.S. Congress mandated that the Department of Education commission a nationwide study to measure, for the first time, the achievements of special education students in the areas of education, employment, and independence. Similarly, in its report The Education of Students with Disabilities: Where Do We Stand? (1989), the National Council on Disability encourages a focus on achieving and assessing advancements in educational quality and student outcomes, rather than a more limited emphasis on the processes and procedures for assuring access to a public education. The reporting requirements of PL 99-457 reflect this shift, as states are being asked to report data on the school leaving status and anticipated service needs of special education exiters.

Perhaps as important as these external mandates in encouraging more outcome assessment is the growing recognition that such assessments can be used to focus institutional attention on critical areas and to improve programs and policies. Although the notion of judging program effectiveness by student achievement and postschool outcomes is somewhat new to special education, school personnel, policymakers, and other stakeholders are quick to recognize the utility and appropriateness of such measures for program improvement.

In response both to this awareness and to the federal mandate, in the past five years or so, several states and school districts have begun to assess special education students' school achievement and obtain follow-up data on their school leavers with disabilities. Results have raised important theoretical questions related to expectations and outcomes of special education, as well as a raft of technical and implementation issues related to the study of these issues.

With this growing interest and activity in outcome assessment in special education, it is time for reflection. What has experience taught us about the strengths and weaknesses of various measures and procedures? Resources for research and evaluation always will be limited; but can we highlight both effective procedures and pitfalls so that we can use resources for outcome assessment to maximum benefit? This report is intended as a positive response to that question.

Our intent is to highlight what has been learned from outcome assessment in special education as a way of improving future research. We draw examples from various outcome assessment projects, with particular emphasis on the National Longitudinal Transition Study of Special Education Students (NLTS), being conducted by SRI International for the Office of Special Education Programs, U.S. Department of Education. This 5-year Congressionally mandated study includes more than 8,000 youth who were ages 13 to 21 and special education students in the 1985-86 school year in more than 300 school districts and 25 state-supported schools nationwide. The NLTS is describing the experiences of youth in all 11 federal disability categories in the

domains of education (both secondary and postsecondary), employment, and personal independence.

In selecting the examples we use, we recognize that outcome assessments are never conducted in a perfect environment. They generally seek to serve multiple purposes for multiple audiences with too few resources and with tools that often are limited or flawed. Further, assessments often are based on information collected by people who have other things to do (e.g., school staff) about people who may not want to cooperate (e.g., school leavers). Not all challenges to good research can be overcome, but their threats to the usefulness of findings of outcome assessments in special education can be minimized if we learn from the experiences of others.

By highlighting "best practices" in special education outcome assessment, we hope to assist those who may be considering or planning outcome assessments in designing such activities in a way that is likely to meet their information goals. By identifying some of the limits of outcome assessment, we hope to assist consumers of such evaluations in interpreting accurately the information they provide.

What Is Outcome Assessment?

Although outcome assessments can differ widely in such key aspects as design, measures, and data collection approaches, they share a common focus on outcomes as individual achievements, statuses, or behaviors. Special education outcomes include those achievements, statuses, or behaviors of special education students that researchers theorize are affected by the educational process. These can include skills or competencies, grades, statuses conferred by the school (e.g., high school graduate), or postschool accomplishments (obtaining employment, enrolling in postsecondary education).

Assessments of such special education outcomes most often have an evaluative purpose in that outcomes measured for special education students (or former students) reflect how well the special education system in general

is doing. Since most students with disabilities receive part of their education in the mainstream, outcome assessment also can describe and evaluate regular education programs. However, even when used for the common purpose of evaluation, outcome assessments can address a wide variety of topics. The purpose can be broad--for example, describing the current employment status of special education graduates at a national, state, or local level. More specific purposes might focus on testing a particular hypothesis (e.g., young people with better social skills are more likely to find and keep competitive jobs) or on examining the effects of a specific intervention or system change (e.g., a change in graduation requirements on the graduation rate of students with learning disabilities in a particular school district).

An outcome assessment's purpose places particular demands or constraints on its design and implementation. For example, in a study that is intended to describe the status of a particular group of young people (e.g., school leavers), it may not be critical to include a comparison group. On the other hand, if the purpose is to determine the effectiveness of a program or policy, the underlying issue often becomes "more effective than what?". This question implies that a comparison will be made and, therefore, necessitates that baseline data are collected or that a control group is specified. For these reasons, the purpose/purposes of a study must be clearly delineated to shape its design.

The Process of Assessing Outcomes

The process of assessing outcomes can be thought of as a sequence of activities, listed in Figure 1. They begin with planning the purposes and procedures of the assessment, continue through data collection and analysis, and conclude with reporting of findings. The remainder of this paper devotes sections to each of the activities in the outcome assessment process, identifying important issues to consider at each step.

Figure 1
KEY ACTIVITIES IN THE OUTCOME ASSESSMENT PROCESS

- Identifying key issues and information needs.
- Developing a conceptual framework to guide the assessment.
- Specifying the nature of comparisons to be made.
- Designing and selecting a sample.
- Selecting and operationalizing outcome measures.
- Choosing independent variables to illustrate outcome variations.
- Selecting data sources and collection methods.
- Choosing analysis methods that are appropriate to the data and to a given project's information needs.
- Communicating findings to encourage their use in policymaking and programming.

Identifying Key Issues and Information Needs

In identifying the issues and information needs to be addressed by an outcome assessment, an emphasis on collaborative planning can help ensure that a study's design is compatible with the information needs of the various stakeholders in the system, the capabilities of collecting and reporting data, and the availability of information. Collaborative planning also can be useful in soliciting support and commitment to an assessment and increasing the likelihood that findings are used appropriately. Consequently, participants should help select the variables to be studied, agree on questions to be addressed, provide input about the design of instruments, and aid in interpreting results and deciding on subsequent plans of action.

- **Collaborative planning increases stakeholders' support and eventual use of outcome assessments and can improve the design of the study.**

Collaborative planning begins by identifying potential contributors to and users of outcome data, while adhering to an organizational structure that facilitates review of information and development of plans. This structure should establish clear linkages among those who develop, manage, and use outcome information and create regular opportunities for interaction.

At no time is input from multiple sources more crucial than in the initial planning stages of outcome assessment, when key issues are identified and information needs are clarified. Informal or formal needs assessment conducted at this stage can serve as the basis for development of a conceptual model, selection and definition of independent and outcome variables, planning data analysis, and structuring timelines and reporting formats. Informal needs assessment can be conducted by forming advisory boards comprised of representatives of key stakeholder groups (i.e., parents, school personnel, students, adult service agency staff). Formal needs assessment might involve systematically sampling and then surveying or interviewing a large group of key stakeholders. Systematic sampling of a large stakeholder group reduces bias that may occur with less formal techniques. It also allows for analysis of findings by stakeholder group, geographic region, or other demographic data of interest. Whether formal or informal, need assessment should address at least the following questions:

- What are the major issues or concerns to be addressed by this outcome assessment?
- What school or program variables, individual, family or community variables, and student outcomes are salient to the above concerns?
- What data sources, existing or planned, are available for use in this effort?
- What capabilities exist among stakeholders to collect, report, and or analyze data?
- What uses exist for the data and what timelines will insure that utility will be maximized?

■ **Needs assessment should be conducted in the earliest stages of planning for outcome assessment and should include all stakeholder groups.**

Though most critical during the initial planning stages, collaboration is necessary throughout the duration of a study. Specifically, participation in the early planning and implementation phases increases the likelihood that stakeholder's interest and needs are represented, bolsters their faith in the findings, and strengthens their commitment to using the findings. Collaboration is equally important when evaluation results are disseminated and used. Program improvement, long-term planning, and needs assessment rarely involve only a single agency or program. To be maximally effective, these processes should reflect the broad context in which a given program operates. Allowing persons from different agencies and different roles who represent different interests to have access to outcome data raises different issues and suggests different solutions depending on the perspective of key stakeholders; further, data interpretation is aided by the insight of multiple perspectives.

- **Representation of multiple perspectives in outcome assessment increases the validity and aids in interpretation of findings.**

The following sections illustrate the subsequent stages in the outcome assessment process, in which key choices can be informed by collaboration, beginning with the development of a conceptual framework.

Developing a Conceptual Framework

An outcome is, by definition, the result of a process. A conceptual framework depicts this process, as well as the relationships between its dynamic and static pieces. As such, the conceptual framework guides the choices to be made at each step in an outcome assessment.

Developing the framework forces the researcher to be explicit from the outset about his or her assumptions regarding what will be measured and why and how data will be analyzed. This step insures that, at the end of the process, findings will meet the information needs they were intended to

serve. Moreover, a conceptual framework provides a structure for understanding, interpreting, and manipulating outcome measures. It answers the question of why a particular outcome is important, and identifies factors that must be taken into account to interpret results appropriately. The conceptual framework is critical to the success of an assessment and should be specified in as much detail as possible.

- **A conceptual framework provides a structure for understanding, interpreting, and manipulating outcomes and should be specified in detail.**

In reviewing 27 follow-up and follow-along studies in special education, Halpern (1987) found that none was based on a conceptual framework that was made explicit by researchers. Despite the recommendation that such outcome assessments "begin with the articulation of a conceptual model that describes the major parameters of the study and guides the development of the research design" (p. 4), many outcome assessments continue to fail to make explicit the conceptual frameworks underlying the approaches they take.

The lack of a concept[illegible] framework can seriously limit the usefulness of the findings of an outcome [illegible]essment. For example, one outcome assessment in an individual state attempted to determine the effectiveness of delivering special education services in regular education placements by comparing regular education students with two groups of special education students: those in regular education placements and those in special education placements. The findings indicated that the school performance of students in special education was poorer than that of both their peers in regular education and nondisabled students. However, the authors acknowledge that the characteristics and abilities of the students in the three groups may have differed greatly and that these differences were not controlled for in the design of the study. Given this limitation, the research could offer no insight about the effectiveness or impact of the different settings--its intended purpose. Use of a conceptual framework would have pointed up the need for additional control variables related to student characteristics and offered hypotheses about what effects differences in student characteristics might have.

Figure 2 presents an example of a conceptual framework that might serve as a guide for an assessment of the impact of secondary special education on postschool outcomes. It illustrates several important aspects of a thoroughly specified conceptual framework.

First, the ultimate outcomes of interest are specified (postschool experiences with employment, postsecondary education, independent living and other productive activities). These distal outcomes are accompanied by specification of intermediate or proximal outcomes (school performance and school completion). Given the complex interaction of individual, family, and community factors that may influence postschool adjustment, it often is difficult to attribute distal outcomes to aspects of school programs or student performance. The inclusion of proximal outcomes is useful for judging the direct impact of education on postschool outcomes. Further, other key independent variables that are expected to influence outcomes are suggested (e.g., individual characteristics), along with the hypothesized path of influence. This type of framework would aid the researcher in obtaining the full range of data needed and in employing an approach that would lead to understanding how school experiences relate to postschool outcomes, one of the intended purposes of the project.

- **Conceptual frameworks should include both proximal and distal outcomes, key independent variables that are expected to influence outcomes, and indications of the expected relationships among them.**

It should be noted that conceptual frameworks can be generic, such as the one specified in Figure 1. Generic frameworks represent commonly held views of educational attainment and contain indicators that research or popular opinion deem important, such as graduation, grades, and so on. It may be that specific frameworks, i.e., those developed for a specific population, such as incarcerated youth, or for a specific purpose, such as an assessment of the effects of minimum competency tests and increased graduation requirements, will vary considerably in terms of outcomes specified, independent variables included, and interactions considered.

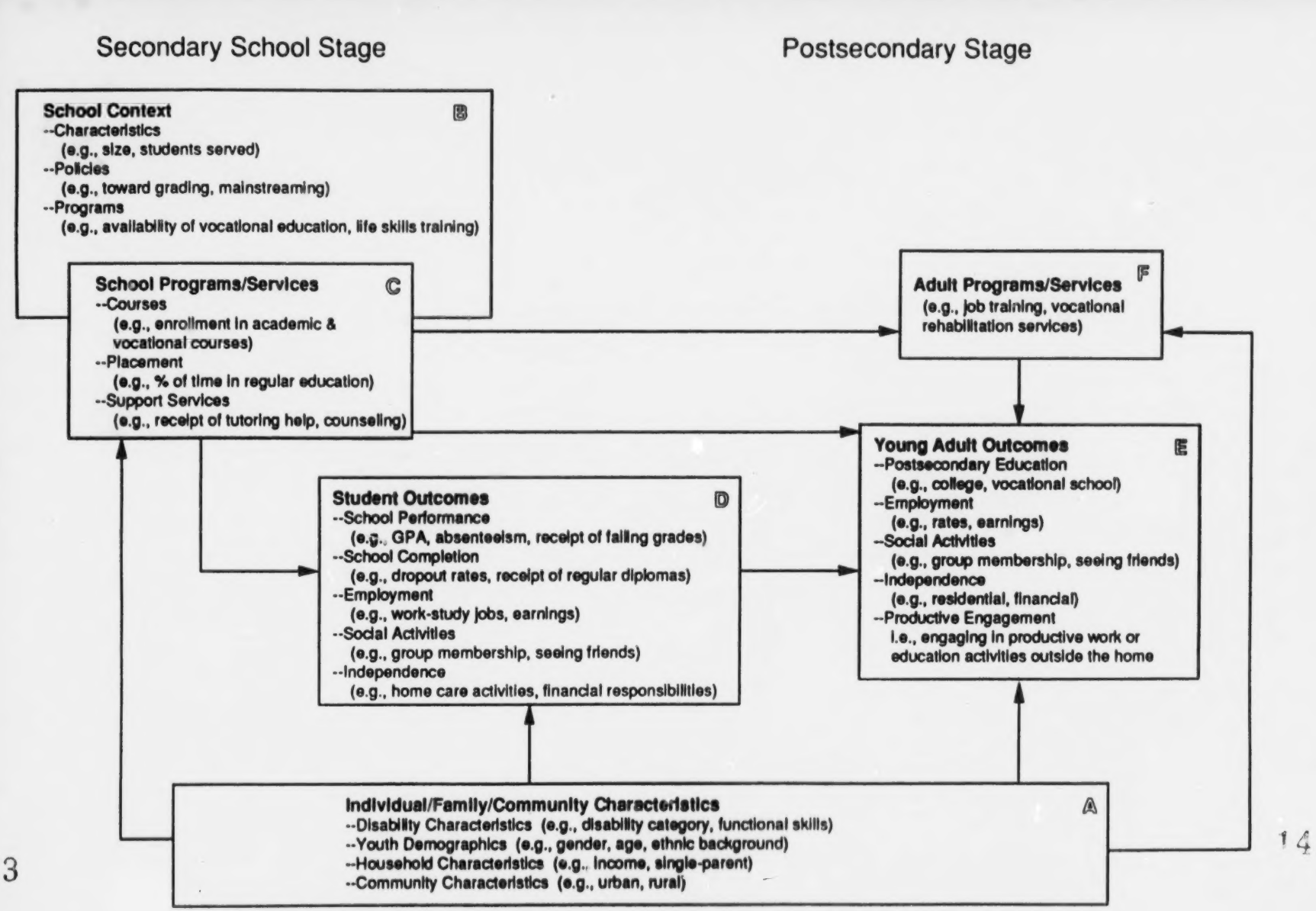


FIGURE 2 CONCEPTUAL FRAMEWORK OF TRANSITION EXPERIENCES AND OUTCOMES OF YOUTH WITH DISABILITIES

- **A conceptual framework can be generic or developed expressly for a specific circumstance.**

Specifying the Nature of Comparisons to Be Made

All outcome assessments imply that data will be used for comparisons. Standing alone, outcome measures do little to inform practitioners, researchers, or policymakers about how well students are doing. For example, we have learned from the NLTS that 32% of youth with disabilities who left secondary school in a two-year period dropped out. It is impossible to determine if that dropout rate is high or low unless we are able to compare it with the dropout rate for another group of young people.

Four common comparisons are used in special education outcome assessments: (a) comparisons with the general population of youth, (b) comparisons among youth in different disability categories, (c) cross-unit comparisons (i.e., cross-school, cross-district, cross-program), and (d) comparisons of the same group over time. Such comparisons are based on the assumption that outcome differences for different groups can be attributed to the factor on which the groups are distinguished (e.g., disability category, exposure to a program). However, for each type of comparison, alternative explanations commonly challenge this attribution. Each type of comparison is discussed below, along with the pitfalls that may limit its usefulness.

- **Comparison groups often are necessary to interpret outcome data, but the validity of comparisons must be carefully assessed.**

Comparisons with the General Population of Youth

Special educators have a continuing interest in understanding the effects of disability on outcomes. One way to determine such effects is to compare the outcomes of young people with disabilities to those of young people from the general population. Generally, differences are assumed to be a result of disability.

Outcomes of the general population can be measured by including a nondisabled control group in an outcome assessment. However, limited research funds and the difficulty associated with securing access to such a group often preclude this approach. Alternatively, extant data on the general population may be used in such comparisons. Census data, High School and Beyond, and the National Longitudinal Survey of Youth are some well known sources of comparison data.

We urge caution in interpreting the results of such comparisons, however. Using data from the NLTS, Figure 3 illustrates that students with disabilities differ from their nondisabled peers in important ways other than disability (Marder and Cox, 1990). Unless these differences in gender, race, urbanicity, income, parental education, and household composition are acknowledged and controlled for, it is impossible to know whether outcome differences are related to the presence of a disability or to demographic differences. In addition to these demographic factors, any attempt to assess the effects of disability on outcomes must take into account mediating factors associated with labelling, such as participation in special school programs, decreased opportunities for interaction with nondisabled peers, or social stigma. It is generally accepted by researchers and advocates that

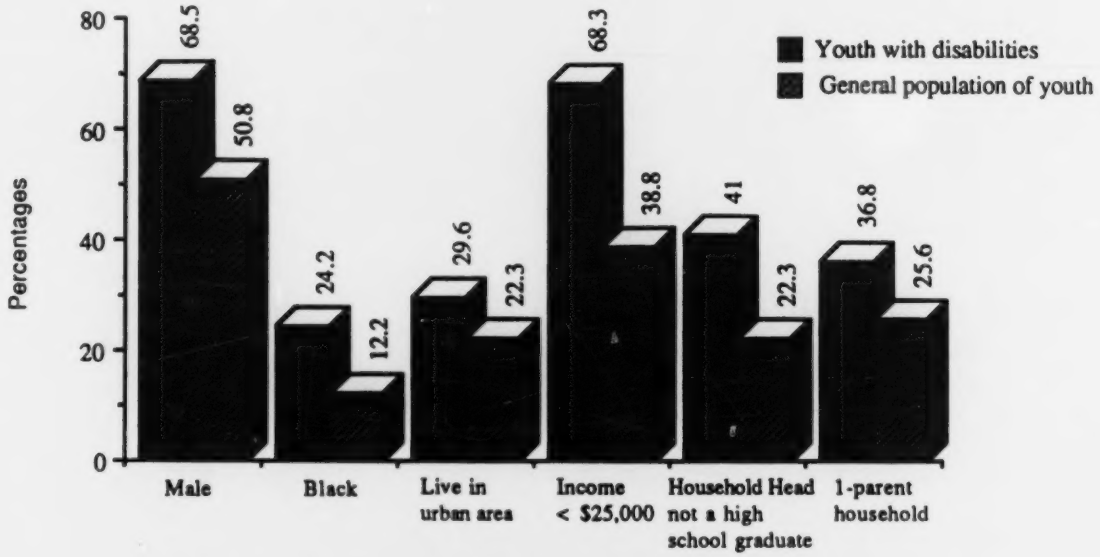


Figure 3: Demographic Differences Between Youth with Disabilities and the General Population of Youth.

these mediating factors associated with disability have substantial impact (sometimes negative) on outcomes. Although such comparisons can provide an important context for understanding student outcomes, researchers must acknowledge alternative explanations for any differences that emerge.

- **When outcomes of youth with disabilities are compared with those of young people from the general population, demographic differences between the two groups should be controlled before differences can be asssumed to related to disability.**

Comparisons of Youth with Different Types of Disabilities

Many outcome assessments in special education, including the NLTS, compare the outcomes of youth in different disability categories in an effort to answer such questions as: How does the school performance of students with sensory impairments differ from that of students with learning disabilities? Do youth with mental retardation achieve competitive employment at a rate different from youth with learning disabilities?

Such comparisons reflect an understanding of the critical influence of the nature of disability on outcomes. Youth in different disability categories can have radically different experiences in school and beyond. In reflecting on this diversity, the NLTS has concluded that:

> In that sense, there is no such thing as 'youth with disabilities as a whole.' In many ways, they differ as much from each other in abilities, disabilities, and experiences as they do from the general population of young people. A (focus on) youth with disabilities...masks this extreme variation and obscures the successes that are apparent. (Wagner, 1990a, p. 11-3).

Hence, disaggregating the population of youth with disabilities by type of disability adds greatly to an understanding of their range of outcomes.

Beyond the differences between disability categories, however, we also know that there is considerable variation among youth who share the same categorical label. Using data from the NLTS, Table 1 demonstrates the

variation in functional skills and IQ that exists within each disability category. When outcome assessments consider disability category only, the variation in abilities and its power to explain differences in outcomes are ignored. For example, a comparison of employment rates of youth categorized as learning disabled with those categorized as mentally retarded might reveal a significantly lower employment rate for youth with mental retardation. An examination of within-category differences, however, might show that students labelled learning disabled who had IQs below 75 and youth labelled mentally retarded with IQs in the same range were similar in their employment experiences, resulting in a more finely tuned and useful understanding of the relationship between disability and employment. Whenever possible an examination of variations in abilities within disability category should be incorporated into the design of outcome assessments.

- **Disaggregating the population of youth with disabilities by type of disability adds greatly to an understanding of outcomes; however variations within disability category should be incorporated into the design of outcome assessments whenever possible.**

Cross-Unit Comparisons

This approach involves comparing outcomes across such units as school districts within a state, schools within a district, or groups of students within a school. Cross-unit comparisons often are used to assess the effectiveness of a particular program, for example, by comparing students in a school in which a program operates and students in a school without the program. To be valid, cross-unit comparisons require giving careful consideration to between-unit differences that may affect outcomes. Demographic differences between students in different settings must be controlled. In addition to demographic differences, different jurisdictions can have different regular education and disability-related philosophies, policies, and practices that may affect outcomes. Such alternative explanations for outcome differences should be explored and made clear to the reader.

- **In cross-unit comparisons, demographic, philosophical, political, and programmatic differences must be accounted for before cross-unit differences can be meaningful.**

Table 1
SELECTED DISABILITY-RELATED CHARACTERISTICS OF YOUTH WITH DISABILITIES

Disability Category	Disability-Related Characteristics: Percentage with High Functional* Mental Skills	N	Percentage with IQ Score: <75	75-90	>90	N
All conditions	56.9 (1.5)	6,585	33.9 (1.6)	41.0 (1.7)	25.1 (1.4)	4,383
Learning disabled	66.0 (2.3)	911	13.6 (1.7)	52.6 (2.6)	33.7 (2.4)	748
Emotionally disturbed	65.3 (2.8)	593	18.4 (2.6)	43.2 (3.3)	38.5 (3.1)	427
Speech impaired	68.9 (3.2)	452	32.3 (4.6)	45.4 (4.9)	22.3 (4.3)	212
Mentally retarded	32.8 (2.2)	860	63.0 (2.1)	16.1 (1.7)	.9 (.4)	803
Visually impaired	31.8 (3.2)	695	25.8 (3.8)	30.4 (4.0)	43.8 (5.0)	465
Hard of hearing	60.7 (3.4)	659	16.3 (3.4)	37.9 (4.7)	45.3 (4.9)	338
Deaf	44.3 (3.1)	743	15.5 (2.5)	28.7 (3.4)	55.8 (4.8)	468
Orthopedically impaired	50.5 (3.5)	628	38.3 (4.1)	41.6 (4.3)	20.1 (3.7)	355
Other health impaired	57.3 (3.7)	411	38.9 (6.2)	30.7 (6.0)	30.3 (6.0)	143
Multiply handicapped	12.8 (2.7)	559	80.7 (3.5)	14.0 (3.2)	5.3 (1.9)	396
Deaf/blind**	6.9 (4.0)	74				

* Parents rated on a 4-point scale youths' abilities (a) to tell time on a clock with hands, (b) look up telephone numbers and use the phone, (c) count change, and (d) read common signs. Ratings were summed to create a scale ranging from 4 to 16. High ability is defined as a scale value of 15 or 16.

** Too few deaf/blind youth had IQ scores to report them separately; they are included among youth with all conditions.

Source: National Longitudinal Transition Study of Special Education Students. Skill scores come from parent interviews, IQ scores from school records from the most recent year in secondary school. Standard errors are in parentheses.

Longitudinal or Time-Series Comparisons

This type of outcome assessment involves repeated measures of the same phenomena taken at several points in time, as a basis for constructing outcome trends. Comparisons of the same group over time can control for demographic or policy differences that plague cross-unit comparisons, but historical influences such as a fluctuating economy, changes in graduation requirements or other policies, and demographic shifts sometimes make attribution of changes observed difficult. For example, instead of reflecting a decline in the ability of high school graduates, the much publicized decrease in SAT scores is largely attributable to a shift over time in the demographic characteristics of the population of students taking the test. Again, researchers are obligated to acknowledge these kinds of alternative explanations for differences in outcomes.

- **Longitudinal or time-series comparisons are affected by historical, economic, and political changes that may confound results.**

Designing and Selecting a Sample

As with any kind of evaluation, the data generated for outcome assessments in special education are only as good as the sample for which they are collected. Weaknesses in sample design are among the most common and most serious threats to the usefulness of findings from outcome assessments. Hence, they are considered in some detail here.

In most outcome studies, it is not necessary or feasible to collect information from every member of a group, especially when the group is large. When it is appropriate or necessary to include only part of a group in a study, a sampling plan must be developed. Below, we discuss five issues that should be addressed in a workable sampling plan for outcome assessments in special education:

- The nature of the population the sample is intended to represent.
- Same size considerations.

- Sample selection methods.
- Problems in locating respondents and acquiring the data.
- The researcher's responsibility to demonstrate generalizability.

What Group(s) Should the Sample Represent?

An obvious first step in selecting a sample for an outcome study is specifying the characteristics and bounds of the target group of individuals with disabilities. For example, if the purpose of a study were to examine the postschool outcomes of special education students in the class of 1988, it would be important to distinguish if that group should include only students who graduated in 1988 or the more heterogeneous group of students who, by virtue of age or class, were supposed to graduate in 1988 but may have dropped out, aged out, or left school by other means at some point in their secondary school years. Comparison groups also should be specified (e.g., a nondisabled comparison group, or students with disabilities who were not exposed to a particular treatment).

Beyond these obvious comparison groups, researchers may wish to stratify the sample by various characteristics of the sampling unit that their conceptual frameworks suggest reflect important differences in the sample. The characteristics that are important will differ according to the purpose of the study and may refer to students (race, gender, handicapping condition), schools/programs (size, instructional strategies, resources), or communities (urbanicity, employment rates, tax base). For example, in assessments of employment, differences between males and females often are found to be large (D'Amico, 1990). If researchers wish to analyze such differences, they may need to stratify the sample by gender to ensure that sufficient cases for both genders are selected. Similarly, if researchers wish to generalize to schools or districts within an entire state, they may wish to stratify a sample of districts by size or urbanicity to ensure that large and small, urban and rural units are represented.

- **Characteristics, bounds, strata, and unit of the target group should be clearly specified before sampling begins.**

Sample Size Considerations

The size of samples that support outcome assessments in special education varies widely. For example, as a national study, the NLTS has gathered data for more than 8,000 youth (Javitz and Wagner, 1990). In the outcome assessments he reviewed, Halpern (1987) found samples ranging from as few as 47 students in one district to more than 1,200 youth sampled across an entire state.

Decisions regarding sample size involve weighing the need for having enough cases to measure outcomes with sufficient precision and to detect significant between-group differences with the costs and complexities of large samples. Although the serious constraint of limited funds is recognized, many outcome assessments are limited in their usefulness because they base conclusions on few cases. An insufficient sample often results from three circumstances: the inability to locate or secure data from those selected for the sample (discussed in the next section), disaggregating the sample into numerous subgroups during analysis, and attrition in the sample over time. The latter two circumstances are discussed here.

Subsetting. Some outcome assessments in special education start out collecting data on a reasonable number of sample members, but in the course of analysis break the sample into ever smaller groups. For example, one study of special education exiters began with a sample of 134 youth, 68 of whom had disabilities and 66 of whom were nondisabled. An analysis of employment segmented each group by gender, yielding samples of 51 males with disabilities and 17 females with disabilities. Four of the females with disabilities were employed. Comparisons of these 4 young women with the 11 employed nondisabled women led the researchers to call for a new federal initiative to address the critical employment problems of young women with disabilities.

The experiences of four young women are an insufficient basis for developing such sweeping policy statements. Although the findings of this project may hold up with larger samples, the confidence in the researchers' conclusion is seriously limited by the small number of cases in their

ultimate analysis, even though their initial sample may have been of reasonable size. Anticipating the subsamples that will be of interest in the analysis is one step toward ensuring a sufficient initial sample to support later analyses.

Attrition. Longitudinal assessments are subject to another reason for ending up with an insufficient sample: attrition over time. When researchers choose a longitudinal design, sample-size estimates should be based on the desired sample at the conclusion of the project, rather than the initial sample. By working backward, therefore, the researcher can increase the initial sample depending on the length of the study. The longer the period of measurement, the larger the initial sample must be to ensure that the sample for which full data are available will support the analyses required.

The NLTS, for example, has experienced a loss of approximately 2% per year of youth who were included in the first wave of data collection in 1987, with higher attrition rates for older youth and those no longer in secondary school. Attrition estimates by other researchers involved with young people range up to 6% per year. Researchers can use such estimates to calculate the initial sample that would be needed to yield the desired concluding sample.

- **Samples must be large enough to measure outcomes with sufficient precision and to detect significant between-group differences. Insufficient sample size is usually attributable to three circumstances: inability to locate or secure data from those selected for the sample, disaggregation during analysis, and attrition. Design considerations can alleviate some of these problems.**

Sample Selection Methods

Samples can be selected in three ways (Wo' then and Sanders, 1987): (1)accessibility--subjects are selected on the basis of physical proximity and willingness to participate; (2) judgment--subjects are selected on the basis of expert opinion or best guesses about who might represent the characteristics of the group; and (3) probability--subjects are selected on the basis of the probability with which they occur in the target group (as a whole or stratified).

Accessibility and judgment as selection strategies may be easy and quick to use, but both strategies are prone to systematic bias and produce samples that may not closely reflect the target population. For example, choosing only districts that volunteer to provide data on student achievement may lead to an overrepresentation of districts with high achievement scores that may be more eager and willing to participate.

Probability samples generally are chosen randomly from a listing of the universe of units that could be included (e.g., all schools implementing a particular program, all students with particular characteristics). Random sampling procedures often are more difficult to accomplish than other strategies by requiring *a priori* identification of and access to members of the target group. However, random sampling increases the likelihood of sample representativeness and should be employed to the greatest extent possible in sample selection.

If using individual students as the sampling unit proves too costly in terms of time and money, or if a list of all members of the population is not obtainable, cluster sampling techniques may be a good alternative. In cluster sampling, the unit of sampling is not the individual but a naturally occurring group of individuals such as classes, schools, or districts. Suppose that one wishes to administer a survey to a random sample of eleventh graders across the state. If random sampling were used, one would obtain a list of all eleventh graders and randomly draw names of individual students. If cluster sampling were used, a listing of all high schools in the state might be obtained, and a random sample of high schools would be chosen. Eleventh graders in the selected high schools would comprise the sample. In a multistage cluster sampling design, once high schools were randomly selected, classrooms within selected schools also would be randomly selected for inclusion in the study.

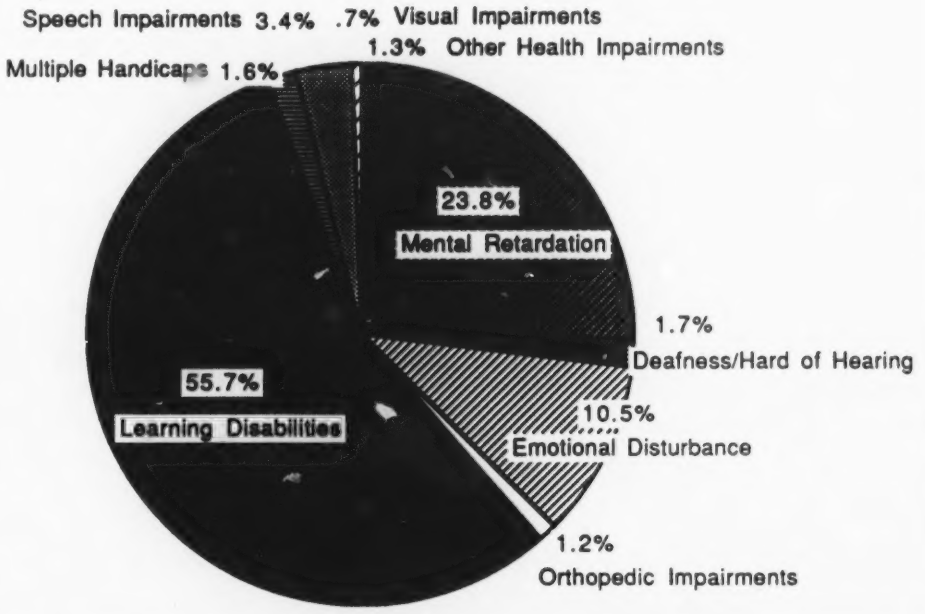
The main advantage of cluster sampling is that it saves time and money. The use of this sampling technique enables one to confine data collection to a small number of sites, making arrangements for access and logistics more manageable. Cluster sampling may be less accurate and less sensitive to population differences than random sampling, but these disadvantages should be weighed against savings in time and money.

The universe from which sample units are randomly selected can be considered as a whole for selection purposes or can be stratified into subgroups (e.g., disability categories, school size), with random selection from the subgroups. For example, if one wanted to draw a sample of students from all students with mental retardation in a school district, one might first identify all students with mild mental retardation, all those with moderate mental retardation, and all those with severe mental retardation in the district. If these three groups were approximately equal in number, a sampling plan could randomly select the same number from each group. If the three groups differed significantly in size, a sampling plan might first determine the percentage of the total group represented in each subgroup, and then randomly select numbers that represent those sample proportions.

- **Sampling can be accomplished by: 1) accessibility; 2) judgement; and 3) probability. Of the three probability sampling increases the likelihood of sample representativeness. Individual, cluster, or stratified random sampling are the most commonly used probability sampling strategies.**

Sampling in special education is sometimes complicated when we wish to measure outcomes of young people with low-incidence disabilities. Figure 4 illustrates this point, using data from the NLTS. At the secondary school

Figure 4: Primary Disability Categories



level, 90% of youth with disabilities are classified aslearning disabled, emotionally disturbed, or mentally retarded. In contrast, youth with sensory, physical, health, or multiple disabilities are very small proportions of the population. If random sampling techniques are used, very few youth with these disabilities are likely to be included in a sample. If researchers want to represent such disabilities, therefore, the universe must be stratified by disability category, along with oversampling of youth with low-incidence conditions.

- **When using stratified sampling, low-incidence conditions should be oversampled.**

The difference between the characteristics of the sample and the characteristics of the population from which the sample was drawn is called sampling error and can be estimated for random samples (Lynch, Hunsburger, 1976). Sampling error is a function of the size of the sample, with error being largest when the sample is small. When probability sampling is used in an outcome study, estimates of sampling error should be presented as part of the findings and used in interpretation.

- **Sampling error should be reported as part of outcome studies.**

Locating Respondents and Obtaining the Data

Many outcome assessments intend to measure outcomes using data gathered from students or families, rather than relying exclusively on data from school records or other extant databases. Obtaining data from students themselves while they are still in school is fairly straightforward; the school constitutes a captive environment, in which students can be observed, interviewed, or tested with relative ease. When parents are chosen as the source of data, or when young people are no longer in school, the difficulty and effort involved in locating sample members is often underestimated.

School records are the most common source of family location information, but they are subject to several weaknesses. For example, the NLTS discovered that some districts do not routinely record parents' names; families are mailed materials addressed to "parents of (name of student)." When trying to locate the family, researchers had considerable difficulty tracking down movers without the name of the parent. Similarly, some districts do not routinely record students' telephone numbers. Yet other districts will not identify special education students to persons outside the school system without written parental consent.

Even when schools do provide full location information, the mobility of many families, particularly in urban areas, makes such information quickly out of date. This problem is exacerbated for researchers attempting to sample young people after they have left school; the longer the time since the students left school, the less accurate the school location information is likely to be.

When the last known address information fails, it is sometimes possible to locate students with moderate and severe disabilities through community service agencies. This system is less successful for students with mild disabilities, who may not access adult services upon leaving school. In retrospective followup studies, students must be pursued through as many sources as possible, including student-friend networks, community colleges and adult education programs, former teachers, and neighborhood canvasses. As a result of the amount of time and effort required to locate students, retrospective studies often are too demanding to be conducted by local districts without external funding or support.

Some of the difficulties associated with following students into adult life are ameliorated by using a prospective approach. Prospective assessment begins systematic data collection and reporting while students are still in school (Edgar, 1988). Before school leaving, permission to maintain contact is obtained from students and parents along with supplementary information, such as students' social security numbers and names and addresses of extended family members or close friends. Upon school leaving, researchers or school

staff maintain periodic telephone or mail contact with students or families. Biannual intervals take advantage of post office and telephone company forwarding procedures. Attrition rates also may be minimized by involving state and local agencies other than the school. Interagency agreements and shared databases among education and vocational rehabilitation, labor, public assistance, and/or mental health agencies can facilitate contact with former students as they move into adult life.

- **Prospective assessment, interagency involvement, and planned cycles of contact can alleviate the difficulties of longitudinal follow-up.**

Unfortunately, the challenges inherent in obtaining data do not end once subjects have been located. Cooperation with data collection efforts also must be secured, although evidence suggests that obtaining cooperation is a much less serious threat to an adequate response rate than the inability to locate sample members. For example, the NLTS demonstrated that almost 30% of students for whom schools provided location information could not be located or interviewed by telephone because the location information was incomplete or inaccurate. In contrast, only 3% of those who were contacted refused to participate in an interview (Wagner, Newman, and Shaver, 1989). A second wave of interviews with out-of-school youth in selected disability categories, however, led to an 8% refusal rate, leading the researchers to speculate that interest in such studies wanes as the temporal distance from secondary school increases.

Regardless of the cause of missing data, failure to obtain an acceptable response rate is a key threat to the accuracy and generalizability of outcome data. Among follow-up studies of special education students, average response rates vary widely, ranging from 27% to 91% (Bruininks and Thurlow, 1988; Schroedel, 1984), depending on the base used in the calculation and the population of young people included. Bruininks and Thurlow (1988) suggest that a 50% response rate is a reasonable expectation for special education students.

Although general factors known to affect response rate, such as method of data collection, survey format, interest in the topic being investigated, follow-up techniques, and use of incentives (Borg and Gall, 1983; Dillman, 1978; Fowler, 1984) appear relevant to these studies, some factors that influence response rates may be unique to samples of students in special education. For example, evidence suggests that the nature and severity of the youths' disability may affect response rates. In a review of 13 follow-up studies in special education, Bruininks, Wolman, and Thurlow (1989) found that studies that followed former students with mild disabilities obtained lower response rates than those surveying persons with moderate, severe, or profound disabilities. This difference may be attributed to the aforementioned problems concerning locating these students, or to motivational factors. Perhaps students with mild disabilities who have left school have been assimilated into the general population and no longer want to be associated with special education. In any case, pilot testing should be done to determine the accuracy of the source of location information, and which survey formats, data collection strategies, and incentives are effective in producing the needed response rate with the specific sample under study.

- **Low response rates are major threats to the accuracy and generalizability of outcome data. Piloting of location information, instrument formats, data collection strategies and incentives can help anticipate and improve response rates.**

Documenting the Generalizability of the Sample

Despite a well-specified sampling plan, some projects end up with a sample that does not represent the group of interest. The factors on which the sample and the target group differ are sources of potential bias in the data if they are related to the outcomes being measured. Researchers must explore issues of bias and present potential sample bias to users of their data. Assessment of bias necessitates determining the comparability of the population the sample purports to represent (e.g., students in the state with mental retardation) and the sample of subjects for whom data are available.

Two factors interrelate in affecting the extent to which bias exists: the percentage of subjects of the total sample selected for whom data were collected (i.e., response rate--Dillman, 1979; Fowler, 1984; Williams and MacDonald, 1986), and the extent to which subjects who were included differ from those the sample is purported to represent.

Theoretically, sample bias is independent of response rate. For example, if a group of 100 students had the same experience (e.g., all were employed), only one student would be needed to represent with accuracy the experiences of the entire group. In reality, however, special education students differ greatly in virtually all dimensions of experience. In our example of 100 students, it may be that 40% were unemployed, 25% were employed competitively part-time, 20% were employed competitively full-time, 10% were employed in sheltered or supported employment, and 5% did volunteer work. If we wish to measure the incidence of various kinds of employment, a majority of the 100 students would need to be measured. When a majority of respondents are successfully included (e.g., 70% or more), issues of bias often are not serious. As the sample proportion declines, however, important aspects of the outcomes are more likely to be missed. Hence, sample bias is often a larger threat as the response rate decreases.

If data are not available for a significant proportion of the sample, it is important to know whether and in what ways the omitted subjects differ from those on whom data were gathered (Dillman, 1978). This can be determined by comparing a common set of data on subjects who responded with data from subjects who did not respond. Bruininks and Thurlow (1988) suggested that for school or postschool studies, school records are a logical source of data on which to make comparisons between respondents and nonrespondents, as they yield data on such characteristics as gender, race, school completion status, grade point average, and absenteeism. For groups with more severe handicaps, comparison may be made on the basis of skill levels or test scores. The NLTS measured bias in a telephone interview sample by conducting in-person interviews with a small subsample; comparisons also were made using school record data (Javitz and Wagner, 1990). Tables showing mean values for the total sample selected and for those on whom data were obtained is a common method for exploring bias.

The presence of bias does not necessarily imply that data are flawed beyond use. Statistical adjustments may be used to correct for differences, although the statistical issues involved in such adjustments can be complex and the assistance of a professional statistician may be needed. Alternatively, data may be interpreted in relation to the group that was represented by the respondents, even when it was not the full group originally intended. For example, if an assessment intended to generalize to youth with the full range of mental retardation, but data were available on few youth with severe retardation, the sample to which data generalize could be redefined as youth with mild or moderate retardation.

Regardless of this choice of handling sample bias, researchers must analyze whether bias exists and state clearly the results of that investigation along with potential effects of bias on their findings.

- **Assessment of bias requires determining the comparability of the population the sample purports to represent and the sample for whom data are available. If bias is found, statistical adjustments may correct for differences or limitations may be placed on interpretations.**

Selecting and Operationalizing Outcome Measures

When looking at outcome assessments, one often can recognize the values and information needs that underlie the choice of outcomes. Many studies chose traditional measures of academic achievement such as grades and standardized test scores as outcome variables. Others represent the outcomes of schooling as combinations of academic and nonacademic skills, such as job, social, or independent living skills. Still others look at students' real-life circumstances (i.e., employment status) as outcomes of schooling. With each choice, the outcomes associated with schooling become further removed from what is traditionally taught in the classroom, thereby extending public education's responsibility beyond the production of literate Americans to the preparation of an independent, productive, and skilled work force. The way in which a program, school, or state views its responsibility will affect the choice of outcomes.

Once the outcome domains are chosen, a further critical choice involves the ways these are operationalized as specific measures or variables. Below, we discuss some of the commonly selected outcomes for students who are still in secondary school and for young people in the postschool period. This discussion of measures has two foci. First, we discuss for each measure common formats or operationalizations of the measures and their various uses. In some cases, we suggest particular definitions to encourage the use of common measures to allow a body of comparable data to accumulate as experience with outcome assessments increases. Second, we discuss the limitations of each measure, recognizing that no perfect measure of outcomes exists. Our point is that measures are often less than they seem, thereby constraining what we can learn from them. When choosing to include a given measure in an assessment of outcomes in special education, researchers must be aware of the implications of their choices and make those implications clear to the users of their research.

Common Measures of Outcomes for Secondary School Students with Disabilities

Grades. Course grades earned by students are common indicators of secondary school performance in studies of student outcomes, both for the general population of students and for students with disabilities (e.g., Donohoe and Zigmond, 1990; Wagner and Shaver, 1989; Wagner, 1990c).

A common operationalization of course grades is a grade point average (GPA), frequently calculated on a 4-point scale by assigning a value of 4 to each "A" grade or equivalent, 3 to each "B", 2 to each "C", 1 to each "D", and no credit to each failed course. Numerical values are summed and divided by the total number of courses completed, including those failed.

An alternative to this operationalization is a dichotomous variable that distinguishes students who received a failing grade from those who passed all courses. Although this second measure of grade performance loses much in the detail of student grade performance, it is useful for distinguishing, in a general way, those students who are "making it" in terms of grades from those who are not meeting the expectations for acceptable performance.

Using grades as outcome measures entails several limitations regardless of the student population involved. For example, the performance level required to earn a particular grade can vary widely from school to school, making cross-school or cross-district comparisons questionable. Further, grade inflation is commonly thought to have eroded the value of grades and pushed up averages, making time-series or longitudinal comparisons questionable. Finally, some jurisdictions employ grading systems that do not lend themselves to calculations of GPAs (e.g., pass/fail) or, in some cases, to any measures of grade performance (e.g., ungraded open education systems).

When we focus attention on grades as outcome measures for special education students some of these limitations become more complex and still others are introduced. Specifically, grade-based measures cannot be calculated for the sizeable fraction of special education students who do not receive grades in their courses. Findings from the NLTS suggest that in their most recent school year, 11% of secondary special education students did not receive grades in any of their courses. As demonstrated in Table 2, an absence of grades is powerfully related to the nature and severity of students' disabilities. Students in some disability categories, students with lower functional skills, and those attending special schools serving only students with disabilities are least likely to receive grades. Hence, using grade-based measures biases the picture of students' grade performance upward relative to what would be found if the performance of all students were measured. Such a bias must be acknowledged by those who select grade-based measures of student outcomes so that users of the information can interpret the findings appropriately.

Further, the meaning of grades for special education students varies depending on whether a course grade was earned in a regular education or a special education class. Data from the NLTS indicate that only 20% of students attended schools that reported using the same grading standard for special education students in regular and special education courses. According to the NLTS (Wagner, 1990c), GPAs are significantly higher (a) for special education courses than regular education courses and (b) for vocational and nonacademic classes than for academic classes. Hence, the GPA

Table 2
STUDENTS WITH DISABILITIES WHO DID NOT RECEIVE COURSE GRADES IN THEIR MOST RECENT SCHOOL YEAR

Student Characteristics	Students Who Did Not Receive Grades: Percentage	Standard Error	N
Total	10.8	1.0	5,591
Primary disability category			
Learning disabled	4.8	1.1	821
Emotionally disturbed	8.7	1.8	502
Speech impaired	4.3	1.5	379
Mentally retarded	24.0	2.0	846
Visually impaired	10.4	2.5	548
Hard of hearing	1.5	1.0	513
Deaf	11.1	2.0	683
Orthopedically impaired	14.9	2.7	458
Other health impaired	9.6	2.6	284
Multiply handicapped	56.1	4.0	491
Deaf/blind	78.1	6.8	66
Functional mental skills*			
Low	54.9	5.3	548
Medium	11.5	1.9	1,724
High	3.6	1.0	1,962
Student attended:			
Special school	54.5	3.9	1529
Regular secondary school	6.9	.8	4052

* Parents rated on a 4-point scale youths' abilities to (a) tell time on a clock with hands, (b) look up telephone numbers and use the phone, (c) count change, and (d) read common signs. Ratings were summed to create a scale ranging from 4 to 16. High ability is defined as a scale value of 15 or 16, medium as a value of 9 through 14, and low as 4 through 8.

Source: National Longitudinal Transition Study of Special Education Students reported in Wagner, 1990c. Grade data are from students' school records, functional abilities data from parent interviews.

for two special education students can vary simply because of differences in the nature and placement of their courses, even when the students' performance is generally at similar levels. These circumstances clearly complicate aggregating grade-based measures for groups of students with different placements. Comparisons of grade-based measures between regular and special education students would be equally confounded by these differences.

- **Grades are commonly used outcome measures. Their use is limited, however, because 1) expectations vary widely making aggregation or comparison difficult; 2) grade inflation limits longitudinal comparisons; and 3) grades are not available for all students.**

Attendance rates. Attendance rates as outcome measures may be used as indicators of a school's or a program's "holding power", that is, its ability to maintain students in a program. This variable is sometimes associated with program factors such as the relevancy of school curriculum, the effectiveness of truancy or other disciplinary or social service programs, or the impact of school policy, such as increased graduation requirements, minimum competency standards, or retention practices. Attendance rates are highly correlated with other outcome variables, such as grade performance and graduation rates (Donohoe and Zigmond, 1990; Schellenberg, Frye, and Tomsic, 1988; Thornton et al., 1987; Wagner and Shaver, 1989; Wagner, 1990b,c).

Attendance measures are usually operationalized as either the number of days or the number of courses for which a student was absent in a given time period. We encourage a consistent use of the number of days absent in operationalizing student attendance because it is the more common metric in school records nationally. It is relatively straightforward to convert a count of courses absent to an equivalent measure of days absent by dividing the number of courses absent by the number of courses students take in a day.

When considering a measure of student attendance for inclusion in an outcome assessment, researchers may face data collection complexities because, in many schools, the files in which attendance data are recorded are separate from students' course-taking and grade records. Hence, using transcripts, for example, as a source of data for school performance may not yield attendance data for a sizeable number of students.

Attendance policies of a school or a district also affect attendance rates. For example, a high school in Illinois implemented a policy whereby parent conferences were held after 10 student absences. During the year following this policy change, the modal number of absences for students in learning disabilities classes dropped from 14 to 9, just enough to avoid the dreaded parent conference.

Interpreting attendance rates as indicators of students' commitment to or involvement in schooling is further complicated for special education students by a prevalence of involuntary absences resulting from health-related aspects of their disabilities. Thus, students with some kinds of disabilities may miss school because of illness or treatments, regardless of their commitment to school. For example, the NLTS found that students in the "other health impaired" and the "emotionally disturbed" categories accumulated the highest average rates of absenteeism of any special education students (16 and 17 days per year, respectively). It is possible, however, that the factors contributing to absenteeism are different for the two groups, making it difficult to infer from such absenteeism data very much about student commitment or school "holding power."

Finally, analyses and reporting of attendance rates must be conducted carefully to avoid misinterpretation. For example, the NLTS analyzed the average number of days absent for a single sample of youth as they aged from 9th through 12th grades; the average number of days absent for this cohort increased each year. However, when separate cohorts of 9th, 10th, 11th, and 12th graders were compared, the average number of days absent declined for each consecutively older cohort. Therefore, depending on the analysis approach selected, two entirely different conclusions would be reached about the attendance trend for students with disabilities across their high school careers. In fact, the single cohort analysis is the more accurate picture of a true attendance trend. The explanation for the second finding likely rests with the fact that the students with the highest absenteeism drop out of school in their earlier grades, thereby purging the older cohorts and contributing to lower absenteeism rates in higher grades.

- **Attendance rates often are used as indicators of schools' "holding power". Definitions and computation of attendance must be uniform across all subgroups in the sample. Analysis of attendance data should consider confounding factors such as attendance policies.**

Suspension. The most common measures of student suspensions are (a) the total number of times a student was suspended over a given time period (e.g., per semester), and (b) the total number of days for which a student

was suspended in a given time period. The measure of incidence indicates the frequency with which behavior problems are severe enough to warrant suspension, whereas the total number of days is a more general indicator of the seriousness of behavior problems (e.g., a single 10-day suspension is counted as equivalent in seriousness to 10 1-day suspensions).

Several measurement issues arise when this outcome measure is selected. For example, because of confidentiality considerations, suspension data often are not reported on school transcripts. Even when recorded in a student's file in a given school year, data related to disciplinary actions are purged from the file in many school districts when a student leaves school. Further, in-house suspensions frequently are not recorded at all.

A further issue arises when one attempts to compare suspension data for different groups of students. Suspension data are affected by the nature of school policy and the consistency with which it is carried out, thereby complicating comparisons of suspension measures across schools or districts. Further, comparison of the suspension rates of special education and regular education students is affected by the fact that 5% of secondary special education students attend schools in which they cannot be suspended (Valdes, Williamson, and Wagner, 1990). Such circumstances reduce the aggregated suspension rate for special education students relative to students in regular education, regardless of differences in behavior.

- **Suspension data, though often included as a measure of frequency or severity of behavior problems, often is not included in or is purged from school records.**

Achievement/competency test scores. Achievement or competency test scores are among the most common outcome measures used for students as a whole, and they are increasingly being used in the context of special education. One difficulty related to using test scores in outcome assessments is the proliferation of tests and the lack of standardization of the grade levels or ages at which tests are given. A lack of comparability of test scores and grade levels makes cross-jurisdictional comparisons particularly difficult.

In the context of special education, test scores suffer from the same "creaming" of students that was discussed relative to grades (i.e., not all special education students are or can be tested). NLTS data suggest that in their most recent school year, 43% of secondary special education students attended schools or were at grade levels for which minimum competency tests were not required. Further, as shown in Table 3, more than one-third of special education students were exempted from such tests, even when they were required of other students. Exemption rates were particularly high for

---

Table 3
SECONDARY SCHOOL STUDENTS WITH DISABILITIES WHO WERE SUBJECT TO MCTS BUT EXEMPTED FROM THE TEST REQUIREMENT

	Students Subject to MCTs Who Were Exempted from the Test		
Student Characteristics	Percentage	Standard Error	N
Total	38.0	2.0	3,325
Primary disability category			
Learning disabled	25.0	3.0	445
Emotionally disturbed	22.2	3.6	273
Speech impaired	12.6	3.1	237
Mentally retarded	72.8	2.6	510
Visually impaired	21.9	3.9	366
Hard of hearing	20.1	3.9	328
Deaf	29.0	3.9	357
Orthopedically impaired	42.0	4.3	303
Other health impaired	23.6	4.6	190
Multiply handicapped	82.7	4.0	288
Deaf/blind	80.0	10.6	28
Student's functional mental skills* were:			
High	25.8	2.9	1,220
Medium	40.0	3.9	1,014
Low	89.0	4.3	335
Student attended:			
Special school	78.5	3.9	861
Regular school	34.2	2.1	2,462

* Parents rated on a 4-point scale youths' abilities to (a) tell time on a clock with hands, (b) look up telephone numbers and use the phone, (c) count change, and (d) read common signs. Ratings were summed to create a scale ranging from 4 to 16. High ability is defined as a scale value of 15 or 16, medium as a value of 9 through 14, and low as 4 through 8.

Source: National Longitudinal Transition Study of Special Education Students reported in Wagner, 1990c; students' school records.

students with mental retardation and multiple handicaps and for students with lower functional skills and who attended special schools serving onlystudents with disabilities. If all special education students were tested, they would register a lower level of competencies overall than would otherwise result from readily available test scores. When researchers use achievement test scores in outcome assessment, they must acknowledge this upward bias in the level of competencies.

- **Lack of comparability of test scores and their invalidity for certain groups limit their usefulness in some outcome assessments.**

School completion status. Students' school completion status has attracted a great deal of attention based on a growing body of evidence suggesting that special education students are disproportionately likely to dropout of school (Butler-Nalin and Padilla, 1989; Mithaug, Martin, Agran, and Rusch, 1988; Wagner, 1990b; Zigmond and Thornton, 1985).

The problems of defining and collecting data on dropout rates have been discussed widely. In addition, the issues involved in defining who is a dropout, the appropriate bases for calculating rates, and the relative merits of event, status, or cohort rates are complex and have been dealt with in detail in other work (Hammack, 1986; 1989; Zigmond and Thornton, 1985b; Edgar, 1988) to which the reader is referred for discussions of the intricacies of calculating dropout rates. Here, we focus on the issues particular to determining school completion status for students with disabilities.

Recognizing the difficulties of defining and calculating dropout rates, it is tempting to focus on what is theoretically its inverse, the graduation rate. This construct offers some advantages over dropout rate in research on the general student population because schools keep relatively reliable records on students when they graduate. Although some differences may exist regarding the definition of school completion or graduation in regular education, these are magnified for special education students. For example, the decision to award a regular diploma, certificate of completion or attendance, or transcript to special education students are often local

ones. Students awarded any of these may be considered graduates, depending on local definitions. For example, in a recent survey of district special education directors, DeStefano and Metzer (in preparation) found that 58% of the districts granted regular diplomas to special education students who had not fulfilled graduation or minimum competency requirements, but who had fulfilled IEP goals; 36% of districts would not grant regular diplomas under such circumstances; and 6% reported that such decisions were made on an individual rather than district basis. These variations in policy make it difficult to compare graduation rates across jurisdictions; besides changes in policy would affect rates over time. An understanding of a given district's policy and the manner in which its graduation rates are computed are important when including this variable in an outcome study.

In addition to dropping out or graduating, students also age out, earn a GED, or enter adult education or alternative programs. The rates at which students pursue these alternative exit routes vary widely for youth in different disability categories and therefore, affect the dropout levels and graduation rates. For example, the NLTS determined that the graduation rates of youth categorized as emotionally disturbed and those categorized as deaf/blind are virtually identical, about 43% (Wagner, 1990b). One might conclude that the school leaving experiences of these two groups, then, are similar. However, a further look illustrates that the most common alternative to graduation for deaf/blind youth is aging out (49%), while virtually half of exiters with emotional disturbances left school by dropping out, a radically different picture. For these reasons, school leaving must be looked at very broadly, including the full range of school-leaving options. When collecting data on school exit status, exit methods must be defined so that respondents report according to common categories.

- **School leaving must be broadly and clearly defined when used as an outcome variable.**

Common Postschool Outcome Measures for Youth with Disabilities

Many outcome assessments choose to look at the postschool status of school leavers (Bruininks, and Thurlow, 1988; Edgar, 1987; Fardig, Algozzine,

Schwartz, Hensel, and Westling; 1985; Hasazi, Gordon, and Roe, 1985; Mithaug, Horiuchi, and Fanning, 1985; Levin, Zigmond, and Birch, 1985; Semmel, Cosden, and Konopak, 1985; Sitlington, 1986; Wehman, Kregel, and Seyfarth, 1985; Zigmond and Thorton, 1985). Commonly collected postsecondary status variables include employment status, postsecondary school enrollment, and residential status. Each variable will be discussed below, along with suggestions for how to expand the range of outcomes examined in such studies.

Employment. In a 1987 review, Halpern found that employment was the most commonly included outcome area in the follow-up and follow-along studies of out-of-school youth with disabilities; 25 of 27 projects measured at least current employment. Although there appears to be some uniformity of interest in employment, considerable variation is apparent in the operationalization of employment measures. Attention to employment can be as simple as a single item asking whether the youth currently has a job, or as complex as requiring a complete work history since the youth left high school. Further, in some studies, employment is defined as paid competitive employment, while in others, employment might include sheltered or supported employment or even voluntary jobs for which youth are not paid. These variations make comparing employment rates across projects or aggregating our knowledge of postschool employment difficult. A more uniform use of at least the following measures would increase our ability to synthesize findings on employment from the many projects considering that outcome area:

- *Current employment status*--whether the youth currently is working in any of the following kinds of jobs: paid competitive, sheltered workshop, supported, or volunteer.
- *Number of hours typically worked per week*--can be collapsed into a dichotomous variable measuring full-time ($\geq$35 hours) or part-time work (<35 hours).
- *Hourly wage*--can be measured directly. Alternatively, hourly wage can be calculated from weekly or monthly earnings, when divided by the number of hours worked per week.
- *Weekly earnings*--Hourly wage alone does not give a sense of an overall level of economic independence (Halpern, 1987). A measure of total earnings for a given time period (we suggest weekly) is needed for that purpose and can be measured directly or calculated by multiplying hours worked per week by the hourly wage.

Another aspect of employment that is frequently measured is job stability, usually operationalized either as the number of months employed at a job (the current job, most recent job, or longest job) or the number of different jobs held in a given time period (the last year, since high school). The proper interpretation for such measures is unclear. Although one might assume that greater stability is a positive outcome, youth just out of high school often need to shift jobs a number of times to acquire skills and experience that enable them to move into positions with career paths. From this perspective, youth having several employment experiences of fairly short duration might be exhibiting a more successful employment pattern than youth holding a single job for a longer period. This ambiguity regarding interpretation leads us to focus on the employment aspects listed above.

- **Employment status can be defined in a number of ways. Multidimensional definitions, including variables related to hours worked, wages, tenure, and satisfaction, permit the clearest understanding of employment as an outcome.**

Postsecondary education. Postsecondary education is a common means for young people to acquire skills and experience for later employment. Although research suggests that youth with disabilities follow this path at a considerably lower rate than the general population of youth (Butler-Nalin and Marder, 1989), outcome assessments can usefully consider postsecondary education measures as adjuncts to employment measures in describing the experiences of youth no longer in secondary school. At a minimum, measures should distinguish whether youth currently are enrolled in each of the following types of postsecondary schools: a vocational or trade school, a 2-year or junior college, a 4-year college or university. Participation since high school also is gathered in some studies. If measures are taken well after secondary school, data can be gathered on whether youth received a degree, certificate, or license from any of the aforementioned kinds of schools. To expand what is learned about this outcome area, measurement might also include the intensity of involvement, in terms of the number of courses taken in a given time period.

- **The measure of postsecondary educational involvement should reflect the nature, duration and intensity of involvement.**

Engagement in productive activities. Employment and postsecondary education are the two most common paths after high school and often are considered separately in assessing postschool outcomes. However, they are not either/or choices. Some youth participate in both. More importantly, some youth participate in neither. Recent research has taken a broader look at postschool outcomes by focusing on the extent to which youth with disabilities became engaged in any of a set of productive activities after high school (Edgar, 1988; Jay, 1990). Alternative conceptions of a measure of engagement have been suggested. The most limited concept measures whether youth worked or attended a postsecondary school (currently or in a given time period). More broadly, job training programs (e.g., being enrolled in a Job Corps program) also can be included. With either measure, gender differences are apparent (Jay, 1990), with young women demonstrating lower levels of engagement. These gender differences are eliminated when a broader definition is used, one that includes being involved in child-raising or other family-care activities. We recommend that engagement be considered more frequently in outcome assessments of youth out of secondary school based on the broadest definition of the concept.

- **Employment and postsecondary education does not entirely define the universe of postschool outcomes for youth. Engagement refers to a broadly defined construct including job training, volunteerism, homemaking, and childcare.**

Residential independence. Although the vast majority of secondary school students live with parents, once youth leave school, residential independence becomes more common, whether or not youth have disabilities (Newman, 1990). Operationally, residential status usually involves assessing the youth's living arrangement (i.e., with parent(s), another family member, or a roommate; alone; or in a hospital/institution, college dormitory, military housing, or a correctional facility). The NLTS considers independence as (a) living alone, (b) with a roommate, (c) in military housing, or (d) in a college dormitory. It is important to interpret residential independence in light of societal trends. For example, economic conditions over the last decade have resulted in larger numbers of youth remaining at home until early adulthood.

Quality of life. Although the concept of quality of life is not new (Flanagan, 1978; Thorndike, 1939), it has recently become an important outcome variable in education and adult services for persons with disabilities for several reasons: technological advances allow it to be measured; research has demonstrated that education can affect it ; complex programs are understood to require complex outcome measures; a growing concern focuses on how persons with disabilities find satisfaction and life quality and how they may be assisted in their efforts to improve it (Baker and Intagliata, 1982; Halpern, Nave, Close, and Nelson, 1986; Hoffman, 1980; Landesman, 1986; Schalock and Lilley, 1986; Zautra and Goodhart, 1979).

Keith, Schalock, and Hoffman (1986) define quality of life as "the degree of independence, productivity, and community integration that a person experiences, as determined by subjective reports or objective evaluations." Subjective measures of quality of life, based on the work of Flanagan (1978), operationalize the dimensions of quality of life on the basis of the perceptions and evaluations of life experience of a large sample. This approach was used by Andrews and Withey (1976), Blair (1977), Baker and Intaglia (1982), and Heal and Chadsey-Rusch (1985).

Objective measures of quality of life, on the other hand, make use of observable, quantifiable indicators of the quality of human experience, such as physical condition, activity level, community involvement, marketable skills learned, mobility, individual decisionmaking, and opportunities for promotion and access to a variety of jobs, living situations, and social interactions (Keith, 1986; Schalock and Keith, 1986).

Because of its multidimensional nature, measuring quality of life requires a significant amount of data collection. Hence, it may not be a feasible component of all outcome assessments. Quality of life is an attractive outcome variable when evaluating programs that attempt to influence directly the independence, productivity, and community integration of a target group, because it represents the broad impact that such interventions can have on several aspects of an individual's life. When program goals are less directly associated with these variables, as in the case of secondary curricula focusing on academics, there is less reason to attribute quality of life status to program effects.

- **Quality of life is an attractive outcome variable for certain outcome assessments. Its multidimensional nature and complexity make measurement difficult, but substantial gains have been made in this area in the last decade.**

Choosing Independent Variables to Illuminate Outcome Variations

As discussed earlier in the section regarding the importance of a conceptual framework, researchers involved in outcome assessments also must make decisions about the independent variables that are necessary to interpret their findings. The overarching purpose of the assessment will affect such choices. For example, an evaluation of particular treatment models will include independent variables that capture important dimensions of that treatment, whereas a broader descriptive look at how youth are doing after secondary school might include independent variables that focus on characteristics of youth. Independent variables often included in outcome assessments focus on students, schools, programs, and communities, such as disability category, program type, urbanicity, or policies that are most likely to influence outcomes.

The importance of demographic factors in explaining outcomes should not be overlooked. The following demographic characteristics can add considerably to an understanding of variations in many kinds of outcomes: gender, ethnicity, age, and household income. Further, some outcome measures can act as independent variables as well. For example, the conceptual framework presented earlier in Figure 2 suggests that school completion is an outcome of school performance, but also a variable that helps explain variations in subsequent postschool outcomes. The conceptual framework developed at the outset of a project is the key to identifying the range of variables needed to illuminate or explain variations in the target outcomes.

Careful thought also should be given to choices of variables describing a program or treatment whose effects are being assessed. Many outcome assessments include single categorical variables describing the major aspect of treatment. For example, if the effects of placement variations were being assessed, a variable might distinguish students who were in regular

education, resource room, or a self-contained placements. If the number of students included in the sample is relatively small, it may be that no further distinctions regarding placement would be possible. However, if sample size permits, an outcome assessment can produce more insightful findings if further aspects of the program or treatment can be measured. For example, one might measure the intensity of a student's exposure to a placement or treatment, such as the percentage of time students were in regular education placements, the number of months over which a student was given tutoring assistance, or an estimate of the total number of hours in a school year that students were provided occupational therapy. With the addition of an intensity variable, students in a treatment could further be categorized as high, medium, or low exposure, if an analysis requiring categorical variables were being employed. Continuous variables could be used in conjunction with variables distinguishing the nature of the program or treatment in many kinds of multivariate analyses.

- **Choices of independent variables should be guided by the conceptual framework that specifies important dimensions of variations in youth or in programs to be considered in an assessment.**

Selecting Data Sources and Collection Methods

Three choices related to data collection have serious implications for the representativeness of the data collected: choice of data source, choice of data collection method, and timing of data collection.

Alternative Data Sources

For some outcome measures, a source of information is readily identifiable. For example, school records are an obvious choice as a source of data for students' course grades. For other outcomes, however, multiple sources make it necessary to select a preferred source.

School records vs. personal reports. Some outcomes, such as school completion status, can be measured using either school records or personal

reports of parents or students. Each source has its own set of limitations. When school records are used as a source of school completion status, for example, a sizeable number of students often cannot be accounted for, as they are reported as "withdrawn," "moved," or "status unknown". Students in these categories accounted for more than 13% of secondary school leavers with disabilities in the 1986-87 school year (U.S. Department of Education, 1989).

NLTS data suggested that when schools were unable to assign a final completion status to students, parents indicated that 62% actually had dropped out. Hence, school records may seriously underestimate the dropout rate. On the other hand, parents may not accurately report completion status. For example, NLTS data suggest that parents may be confused by what constitutes graduation from high school; 60% of parents whose children's school records indicated they had "aged out" reported that the children had graduated. Relying on parents for data on school completion may overestimate graduation rates.

Parents also may be confused about the kinds of services their children receive, suggesting that records may be a more reliable source of such information. For example, the NLTS asked parents whether their children had ever received "training in job skills, career counseling, help in finding a job, or any other vocational education" Researchers found that 62% of youth whose parents responded "no" to that question had taken at least one vocational education course in their most recent year in secondary school (Wagner and Javitz, in process). In this case, parents would seriously underestimate the extent to which youth had received vocational services.

The choice of data source is often constrained by considerations other than data accuracy. For example, limited resources may prohibit researchers from accessing school records as an additional source of data about services when the primary data source is parent interviews. Or access to records may require obtaining written parental consent, which can be time consuming and often unsuccessful. Regardless of the choice of data source, researchers are obligated to identify the sources they use, to be aware of the limitations inherent in their choices, and to state those limitations clearly for the users of their data.

Parents/adults vs. youth. When personal reports are selected as a data source, the choice of respondent becomes an important issue. Many studies have concluded that students with mild disabilities can serve as accurate and reliable informants about their own experiences (Bruininks and Thurlow, 1988; Hasazi et al., 1985; Zigmond and Thornton, 1985). In contrast, accuracy and reliability come more into question as severity of disability increases. In the NLTS, parents were asked whether they believed their children with disabilities could respond to interview questions for themselves. As shown in Table 4, the percentage of parents who reported that their children could be interviewed declined sharply as children's functional abilities and IQ decreased. Therefore, when youth are selected as the respondent, researchers must recognize that they are obtaining data from the most capable youth in a given disability category and that the results will be biased accordingly.

When dealing with students who are young enough still to be in secondary school (and generally living at home with parents) or young people who have moderate and severe disabilities and are not capable of responding to survey questions, it is generally accepted to use parents or other knowledgeable adults as respondents. Acceptability is less clear in cases where students with mild disabilities are unavailable or unwilling to respond, and parents' reports are consequently substituted for youths' responses. Parent and youth responses may differ (Freeman and Medoff, 1982). For example, in an attempt to determine the reliability between parents' and youths' responses on a follow-up survey, Edgar (personal correspondence, 1989) failed to find 100% agreement on any variable, even sex of the youth.

Numerous factors must be considered when determining the appropriateness of parents as respondents. First, the extent to which the parent has contact with the youth is an important consideration. If the youth still resides at home, parents may be aware of work schedules, wages, and social activities. If not, parent reports may be based on general impressions rather than direct knowledge of their child's status.

In an effort to assess how knowledgeable parents were about youth who were no longer in secondary school, the NLTS asked parents how often they saw

Table 4
VARIATIONS IN PARENTS' REPORTS OF WHETHER YOUTH WITH DISABILITIES COULD BE INTERVIEWED BY TELEPHONE

Youth Characteristics	Parents Reporting Youth Could Be Interviewed: Percentage*	N
Total	71.7	6,538
Primary disability category		
Learning disabled	95.9	911
Emotionally disturbed	91.9	590
Speech impaired	91.3	449
Mentally retarded	60.6	840
Visually impaired	90.2	713
Hard of hearing	72.8	647
Deaf	34.1	746
Orthopedically impaired	83.2	613
Other health impaired	77.2	403
Multiply handicapped	29.6	548
Deaf/blind	3.8	78
Self-care abilities**		
High (11 or 12)	80.4	5,020
Medium (7 to 10)	50.2	874
Low (3 to 6)	23.7	514
Functional mental skills***		
High (15 or 16)	90.9	3,052
Medium (9 to 14)	68.3	2,226
Low (4 to 8)	24.8	1,056
IQ score		
85 or more	82.6	1,306
71 to 85	82.6	949
52 to 70	65.6	529
Below 52	25.7	405

* Percentages are unweighted.

** Parents rated on a 4-point scale youths' abilities to dress themselves, feed themselves, and get around to nearby places outside the home. Ratings were summed to create a scale ranging from 3 to 12.

*** Parents rated on a 4-point scale youths' abilities to tell time on a clock with hands, look up telephone numbers and use the phone, count change, and read common signs. Ratings were summed to create a scale ranging from 4 to 16.

Source: National Longitudinal Transition Study of Special Education Students. Skill scores come from parent interviews, IQ scores from school records from youth's most recent year in secondary school.

or talked to their children. As shown in Table 5, the vast majority reported quite frequent contacts with their children, providing some reassurance as to their knowledge of their children's experiences.

Second, the appropriateness of parents' responses may be related to the information requested. For example, although parents may be able to report accurately whether or not their child is employed, they may not know as accurately his/her hourly wage, hours worked, or possibility of promotion. Parents are clearly inappropriate as respondents for items related to such variables as the youth's satisfaction with his/her job or other issues based on attitudes or perceptions, where young people are the only acceptable respondents. In any case, researchers must clearly identify the data source. In addition, it must be specified when data from youth and parents are combined.

- **Data sources should be pilot-tested to determine availability, access, and ease of data collection. When multiple data sources are used findings should be clearly attributed to source. Limitations of data sources should be acknowledged in all reporting and the impact of these limitations on the quality of data should be considered during interpretation.**

Table 5
FREQUENCY OF PARENT CONTACT WITH OUT-OF-SCHOOL YOUTH WITH DISABILITIES

Frequency of Contact	Percentage
Youth lives at home (assumes daily contact)	56.3
Almost every day	11.2
A few times per week	12.6
Once a week	9.8
Every few weeks	7.1
Every few months or less	3.0
N	813*

Source: National Longitudinal Transition Study of Special Education Students: parent interviews.

* Youth were out of secondary school 2 to 4 years, did not live with parents and were classified as learning disabled, emotionally disturbed, speech impaired, or mildly or moderately mentally retarded.

Data Collection Methods

When personal reports are selected as a data source, three collection methods may be employed: self-administered written questionnaires (often mailed to respondents), telephone interviews, and in-person interviews. In selecting among these options, several considerations must be weighed; in some cases, more than one method may be employed.

The nature of the data sought greatly affects the data collection method. For example, if researchers are interested in outcomes measured in the respondents' terms, rather than in prespecified categories, written questionnaires are not recommended. Respondents are rarely interested in or competent to write detailed, open-ended responses.

Costs also must be considered. It is much less expensive to mail questionnaires than to do either form of interview, but considerable effort often is required to achieve an acceptable response rate. Multiple repeat mailings and reminder telephone calls often are necessary, boosting the costs of such an approach.

The nature of the sample also has implications for choosing a data collection method. If a sample is distributed across an entire state, for example, in-person interviews may not be feasible. Alternatively, if the sample contains a substantial percentage of low-income households, a telephone approach can introduce a significant bias in the data collected. For example, the NLTS determined that its sample of youth for whom data were collected by telephone significantly underrepresented low-income, minority households when compared to a sample of nonrespondents to the telephone interview that were subsequently interviewed in person (Javitz and Wagner, 1990). Statistical adjustments were needed to eliminate this bias. As argued in an earlier section on the generalizability of samples, researchers are obligated to demonstrate the extent to which the data produced through the chosen collection methods represent the population intended.

Researchers may want to consider the creative use of a variety of data collection approaches. For example, although the NLTS relied heavily on

telephone interviews, brief written questionnaires were mailed to respondents for whom no telephone numbers were available. Written questionnaires also will be employed in later stages of the study to solicit information from deaf youth who do not participate in telephone interviews. In-person interviews also were conducted to supplement the telephone interview sample in areas with a high rate of nonresponse.

- **The nature of the data, costs, and nature of the sample influence the method of data collection. Record review, questionnaires, and interviews are the most commonly used data collection methods in outcome assessment.**

Instrument Development

In some cases, it may be necessary to develop original instruments for surveys or interviews. The specifics of instrument development are often dictated by the particular context of the research and are beyond the scope of this paper, but some general guidelines may be offered. First, development of reliable and valid instruments can be costly and time-consuming and may be beyond the capabilities of a time-limited project and its professional staff. The possibility of adopting or adapting existing instrumentation used by related projects should be thoroughly investigated before original instrument development is considered. Instrumentation used in the NLTS, High School and Beyond, and many of the other studies mentioned in this paper are available at little or no cost and are applicable to a large number of outcome assessments in special education.

Second, even if previously-used instruments are adopted, pilot-testing must be done to test the appropriateness and clarity of the questions asked for the actual persons who will be responding and to make sure that the response format (i.e., verbal, written, pointing) is appropriate for the group. To fulfill both these purposes, pilot testing must be done with a sample virtually identical to the sample to be used in the study. In addition to testing items and format, pilot tests can be used to refine data collection procedures, to estimate response rates, and to provide a pilot data set. This data set can be used to inform planning for data analysis in terms of estimating the underlying distribution of variables, percentages of missing data, and sample size.

Timing of Data Collection

When studying postsecondary outcomes of students in special education, for example, it is necessary to determine how much time should elapse between school leaving and measurement of outcomes. As this interval increases, it becomes increasingly difficult to attribute outcomes to the effects of schooling without controlling for numerous other factors that may intervene over time, such as fluctuations in the labor market, participation in additional training, or changes in health or family status.

Data collection also becomes more complicated as time elapses; records can be lost, persons may be more difficult to find; perceptions of school and the ability to reconstruct past events may erode; and refusals to cooperate may increase. Such considerations argue for measuring outcomes, at least the first time, fairly soon after school leaving (perhaps six months or a year), a strategy employed by many follow-up/follow-along studies.

On the other hand, it takes time to establish oneself as an adult, making it very unlikely that the postschool status six months after school leaving is indicative of the later postschool status. D'Amico (1990), for example, found that employment rates for out-of-school youth with learning, speech, or emotional disabilities or mild/moderate mental retardation increased steadily in the first four years after high school. Establishing multiple points of data collection at yearly intervals after leaving school can capture such fluctuations or trends, while at the same time requiring the respondent to recall only the last year to allow for a more accurate depiction of each time period.

- **The timing of data collection should optimize the availability and validity of data to be collected.**

Choosing Analysis Methods

Decisions about how data will be analyzed should be made in the early planning stages of an outcome assessment, in conjunction with decisions about information needs, variables and their measurement, data sources, and

audiences. Planning early for data collection prevents the unfortunate circumstance faced by some researchers, who after completing data collection, find themselves unable to answer key questions because of analytic shortcomings such as insufficient sample size, missing data, inadequate level of measurement, or large measurement errors. Planning for data analysis can be facilitated by the use of a management plan or planning matrix as shown in Figure 5. In this plan, major factors to be considered in data analysis are systematically addressed and cautions, concerns and questions are noted. The analysis plan can evolve as data collection begins and more information is discovered about the availability and quality of data. Establishing such a plan early allows for anticipation of major problems and increases the likelihood that meaningful findings will be produced.

As shown in Figure 5, planning for analysis requires the consideration of a number of factors including: a) the nature of the questions asked; b) the characteristics of important variables; c) sample size and composition; and d) the knowledge base and experiences of the audiences who will receive the results of the analysis. It is also wise to anticipate the personnel, time, and technical resources that the analysis may require. As each factor is considered, questions and concerns might be noted for further thought or investigation. A discussion of each of these follows.

The Nature of Research Questions Asked

The evaluation questions or hypotheses should provide the first clue as to what an appropriate data analysis strategy might be. The question, if well stated, should specify the sample and comparison groups, independent and dependent variables and the relationships of interest between them. In Figure 5, the question "Are there differences in basic skill acquisition (as evidenced by gain scores on reading and math standardized tests) attributable to differences in model of special education service delivery (regular education, resource room, self-contained) for a group of eleventh grade students with learning disabilities?" the independent variable, model of special education service delivery, is used to define three comparison

Figure 5: Planning Matrix for Data Analysis

Evaluation/Research Question	Variables/Instrumentation	Level of Measurement/ Underlying Distribution	Sample Size/ Comparison Groups	Candidate Analysis and Underlying Assumptions	Knowledge Base and Experiences of Audiences	Personnel and Resources Necessary for Data Analysis
Are there differences in basic skill acquisition (as evidenced by gain scores on reading and math standardized tests) attributable to differences in model of special education service delivery (regular education, resource room, self-contained) for a group of eleventh grade students with learning disabilities? (Comparison of gain scores across models of service delivery)	Independent variable: Model of special education service delivery a) regular education b) resource room c) self contained Program descriptions will be collected in January, 1991 using a form designed for this purpose. Dependent variables: Basic skill acquisition 1) math gain scores 2) reading gain scores Continuous; interval/gain scores obtained by subtracting ninth grade raw score from eleventh grade raw score. All programs will test students using the ITBS during May 1991. Raw scores on math and reading subtests will be available in June. *(How were subjects assigned to model of special education service delivery?)	Nominal; balanced design equal number of students in each of 3 levels (n=500; N=1500). Gain scores normally distributed (Are variances of three groups equal?)	$n^1=n^2=n^3 = 500$, $n = 500$ Due to attrition, mobility, and exemption/absence from testing gain scores are available for about 60% of the sample. *(Representativeness of remaining scores??)	Analysis: One way analysis of variance comparing average gain scores across the three groups. Post-hoc pairwise comparisons if overall ANOVA findings are significant. Assumptions: 1) homogeneity of variances 2) underlying normal distribution of dependent variable *(How did groups compare on pretest?) *(Within-group differences or uncontrolled between-group differences may be confounding.) *(Analysis of covariance may be used if evidence of pretreatment differences exist.)	Primary audiences: Special education policy makers: may not have an understanding of ANOVA. Graphic displays and narrative necessary.	Local sites will report data via modem using prearranged coding schedule. Data analysis will be performed using SAS-PC. A subcontract will be issued for programming and report generation associated with data analysis. *(Who could do the actual programming for data analysis?)

* Indicate questions or issues to be considered

BEST COPY AVAILABLE

groups: regular education, resource room, and self-contained for a sample of eleventh graders with learning disabilities. It is also implied that a comparison of gain scores in math and reading for these three groups will be necessary to answer the question. Therefore, statistical analysis techniques that can accommodate comparisons of three groups are warranted. Other types of questions may address relationships among variables that are not comparative in nature. Figure 6, adapted from Kleinbaum, Kupper, and Muller (1988) illustrates the relationship among the general purpose of the analysis, as derived from the research question, and the type of data analysis used.

The Characteristics of Important Variables

As seen in Figure 6, the level of measurement of independent and dependent variables is an important factor in determining the type of data analysis that can be done. There are four levels of measurement:

- *Nominal measurement* is the lowest level of measurement and has the fewest analytic options. At this level, values of a variable simply indicate different categories. The variable "gender" is nominal with two values, "male" and "female".

- *Ordinal measurement* allows grouping into categories as well as ordering of the categories. Grades can be thought of as ordinal variables. In this system an ordering can be made of categories, but little information is available on the magnitude of differences between categories.

- *Interval variables* order categories and give a meaningful measure of the distance between categories. Test scores are often considered to be interval data. Interval variables are usually continuous, that is, they may take on any value within a specified range.

- *Ratio variables* represent the highest level of measurement and possess all the characteristics of interval variables in addition to having a meaningful zero point. Physical measurements such as some temperature scales and measures of height and weight are examples of ratio scales. There are a few educational outcome variables that can be expressed on a ratio scale.

More complete discussions of levels of measurement and their impact on analysis are available elsewhere (see for example Kleinbaum, Kupper, and Muller, 1988). For our purposes, it suffices to say that level of

Figure 6. Rough Guide to Data Analysis (Adapted from Kleinbaum, Kupper, & Muller, 1988).

Purpose of the Analysis	Level of Measurement/ Type of Variable		Type of Analysis
	Dependent	Independent	
To describe the general characteristics of a group, a subgroup or a series of groups	Variables may be of any type		Measures of central tendency and variability; frequency distributions
To describe the relationship between two or more nominal variables	Nominal	Nominal	Chi-square or other non-parametric techniques
To describe the extent, direction, and strength of the relationship between several independent variables and a continuous dependent variable	Continuous	Classically all conntinuous, but in practice any type(s) can be used	Multiple regression analysis
To describe the relationship between a continuous dependent variable and one or more nominal independent variables	Continuous	All nominal	Analysis of variance
To describe the relationship between a continuous dependent variable and one or more nominal independent variables, controlling for the effect of one or more continuous independent variables	Continuous	Mixture of nominal variables and continuous variables (the latter used as control variables)	Analysis of covariance
To determine how one or more independent variables can be used to discriminate among different categories of a nominal dependent variable	Nominal	Classically all continuous, but in practice a mixture of various types can be used as long as some are continuous	Discriminant analysis
To define one or more new composite variables called factors from other, specifically constructed or reduced variables	(The variables used in a factor analysis are classically continuous, but in practice may be of any type. These variables are not clearly identifiable as either dependent or independent, although the resulting factors may be used as dependent or independent variables in a later analysis.)		Factor analysis
To describe the relationship between a nominal dependent variable and several nominal or ordinal independent variables, although applications to situations involving only dependent variables are possible	Nominal	Mostly nominal, but sometimes ordinal	Categorical data analysis using linear models

measurement should be considered for each variable when planning measurement and data collection. The design should strive for the highest level of measurement possible for each variable, thus increasing the number of possible analysis options. In analysis, the same variable may be considered at one level of measurement in one analysis and at a different level in another. For example, age may be considered as an interval in a regression analysis or, by being grouped into categories, nominal in analysis of variance.

In addition to level of measurement, the underlying distribution of the dependent variable is sometimes a consideration when selecting an analysis approach. Some types of analysis, like analysis of variance, assume that the dependent variable is approximately normally distributed. This assumption should be tested using pilot data or previous research. Technical consultants or a good statistics book can be invaluable at this stage of planning the analysis.

In Figure 5, it appears that the independent variable is nominal in nature, while the two dependent variables are interval with an underlying normal distribution. According to Figure 6, information suggests that analysis of variance may be an appropriate analysis technique to use. This consideration raises two issues (denoted by asterisks and parentheses). Because analysis of variance assumes random assignment to groups and homogeneity of variance: -- how were the groups formed, and are the variances of the groups equal? These questions can be answered through exploratory data analysis or review of procedures. If basic assumptions are violated, the analysis may still be appropriate, but caution must be taken in interpretation.

Sample Size and Composition

Sample size and the presence of comparison groups may affect the choice of analysis. It should be noted that sample size must be considered at the individual variable level. For example, if a subgroup contains 30 students, one may assume that there are 30 observations for each variable. However, it

may be that attendance information was missing for school records for 11 students in this group. This missing data reduces the sample size of the group to 19 in analyses using the attendance variable.

Some types of analyses such as factor analysis, require large sample sizes. Once again technical consultation can help resolve sample size issues.

Even when sample size is not greatly affected by missing data, there is a concern that bias may be present; i.e., that the group for whom data are missing may differ from the remaining group in important ways. In the example in Figure 5, test scores for 40% of the sample are missing. This calls into question the representativeness of the remaining sample and advocates caution when interpreting this finding.

Knowledge Base and Experiences of Audiences

The knowledge base and experiences of the audience should not dictate the choice of analysis, per se, but should be considered when deciding how to report findings and disseminate results. For example, in Figure 6 the primary audiences may not be comfortable with interpreting an ANOVA summary table. Narrative and graphics displays may be necessary to enhance their ability to use findings.

In the example given in Figure 6, based on consideration of all the factors presented in the matrix, a one-way analysis of variance was chosen as the primary analysis to be used to answer this research question. Confounding factors such as pretreatment or within group differences are noted and the suggestion is made to consider an analysis of covariance if these confounding factors prove problematic. A plan such as this one, produced early in the planning stage of an outcome assessment and revised throughout the implementation stage, can serve as a useful guide. It's use continues through the final stages of the project, when communicating outcome information.

- **Planning for data analysis should occur early in the design phase of the project.**

Communicating Outcome Information

Outcome assessments pay off when decision-makers use the results to affect policies and programs. Yet, serious obstacles often impede such use. Obstacles can be minimized by collaborative planning throughout the outcome assessment process. Use of findings also will be encouraged when findings are based on valid, reliable data. Outcome assessments, in turn, increase their chances of producing such data when they are based on a solid conceptual framework and attend to the methodology issues discussed here. Even then, however, barriers to appropriate use can arise from the ways in which findings are communicated. Such barriers include (a) organization of findings around data rather than issues, (b) limited interpretation of the meaning of data, and (c) reliance on excessively bulky reports.

Many reports of outcome assessments and other studies that rely heavily on quantitative data often focus primarily on those data, providing abundant text and even more abundant data tables. Practitioners and policymakers often have little interest in the data per se. Instead, they turn to outcome assessment with a question or a series of questions, and are primarily interested in the answers to those questions as suggested by the data. As a result, findings must focus on the questions and their answers, rather than on the data themselves.

Similarly, reports of outcome assessments often describe in detail sampling, data-collection, and analysis procedures before outlining the results. Even when the findings are presented, their meaning is not always clear. Often researchers present what they did and what they found, without reporting what they learned. For meaning to emerge, the data must be interpreted, not just presented.

Using an example from an earlier section in this report, we may find that 32% of school leavers with disabilities left school by dropping out, but

what does that mean? When we compare this dropout rate to that of non-disabled students (estimated to be about 25%), we learn that students with disabilities are disproportionately likely to leave school without the skills and credentials implied by a high school diploma. Therefore, they are disproportionately likely to suffer the poor economic consequences that may accompany a lack of skills. The data suggest that schools might usefully focus dropout prevention strategies and resources on identifying and helping students with disabilities. If analyses examined variations in dropout rates for youth in different disability categories, researchers could recommend that such efforts should particularly target youth with learning disabilities and emotional/behavioral disorders as those groups are most likely to leave school early. If analyses also examined variations by demographic characteristics of youth, other factors associated with being at risk of early school leaving could be identified.

As this example demonstrates, the meaning of findings becomes clear when data are interpreted, not just presented. When practitioners and policymakers are not in a good position to interpret data for themselves, it is the responsibility of the researcher to do so. On the other hand, a strong caution needs to be made not to exceed the bounds of the data when making recommendations.

Direct relationships should be evident between the findings of an outcome assessment and recommendations that are made. Further, limitations of the outcome assessment such as nonrepresentativeness of the sample should be clearly stated. Small sample size, measurement error or other threats to validity should be openly acknowledged in any presentation of findings and should figure prominently when deciding what recommendations to make. High-stakes decisions, such as those affecting policy and programs, should be Well-grounded in high quality, verifiable data.

Finally, the findings of most outcome assessments are presented in the form of a "final report." Such a report can be a useful vehicle for summarizing in a single document what was done, what was found, and what was learned. However, in summarizing this breadth of information, reports can be

lengthy and technical. Even when accompanied by a brief "executive summary," final reports are rarely a format that encourages use of the information they contain.

Instead, findings of an outcome assessment are best presented in forms that acknowledges that there are multiple audiences with multiple interests in and uses for those findings. No single format (e.g., a final report) is likely to meet those multiple interests. Alternative dissemination strategies include journal articles, which may best reach other researchers. Relatively brief reports on the outcomes of youth in individual disability categories are being prepared from one outcome assessment project, recognizing that many practitioners specialize in serving youth with a particular disability (Sitlington, 1989). Single-page "highlights" are being produced by the NLTS, each of which focuses on a particular issue (e.g., dropout behavior) or a particular disability category. These publications may satisfy the information needs of practitioners who want a brief summary of the "bottom line" relative to the issue or type of student their programs address. Use of findings is facilitated when they are packaged in a variety of forms and disseminated through a variety of channels.

- **Reports on outcome assessments should 1) be available in several forms, 2) be organized around issues rather than data, 3) be concise, and 4) provide interpretation when necessary.**

Outcome Information in Use: Opening Pandora's Box

Outcome assessments can respond to information needs with valid and reliable data collected from an appropriate sample in appropriate ways and presented in appropriate formats that facilitate their use. Our purpose has been to recommend ways to ensure maximum use and benefit of findings.

When an outcome assessment reports findings that fulfill the project's purposes, researchers may find that rather than being completed, their job has just begun. Good information can be addictive. Good information about outcomes in special education can have a powerful effect on policies and

programming in ways that make the need for continued or further information even more important. Some outcome assessments may point up areas of critical need for program initiatives. If acted upon, such initiatives may necessitate further information on outcomes to assess whether the initiatives are having their intended (and/or some unintended) effects. If trends in outcomes are plotted in an assessment, decisionmakers may wonder in what ways they will fluctuate as we move into the future. In such cases, outcome assessment may evolve from being thought of as a special project to becoming a routine part of planning and programming. As this evolution occurs, new questions arise:

- Who is responsible for producing the outcome data?
- What is the process for revising or redirecting the focus of outcome assessment as new issues or questions arise?
- Where will the resources for routine outcome assessment come from?
- How much information is enough?

Educators and policymakers in several states and communities are grappling with these questions as they seek to incorporate the results of special education outcome assessment into their standard operating procedures. Although we cannot know the right answers for their individual cases, we can support them in their questioning. In recognizing the value of information about outcomes, they are helping the field of special education move toward more effective policies and programs for young people with disabilities.

REFERENCES

Andrews, F.R., and Withey, S.B. (1976). Social indicators of well-being: Americans' perceptions of life quality. New York: Plenum Press.

Baker, F., and Intagliata, J. (1982). Quality of life in the evaluation of community support services. Evaluation and Program Planning, 5, 69-79.

Blair, T.H. (1977). Quality of life, social indicators, and criteria of change. Professional Psychology, 8(4), 464-473.

Borg, W.R., and Gall, M.D. (1983). Educational research: An introduction. New York: Longman.

Bruininks, R.H., Lewis, D.R., and Thurlow, M.L. (Eds.) (1988). Assessing outcome, costs, and benefits of special education programs (Project Report Number 88-1. Minneapolis, MN: University of Minnesota, University Affiliated Program.

Bruininks, R.H., and Thurlow, M.L. (1988). Evaluating post-school transition of secondary students with moderate to severe handicaps (Final Report). Minneapolis, MN: University of Minnesota, University Affiliated Program.

Butler-Nalin, P. and Marder, C. (1989). Making the transition: An explanatory model of special education students' participation in postsecondary education. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Butler-Nalin, P. and Padilla, C. (1989). Dropouts: The relationship of students characteristics, behaviors and performance for special education students. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Colton, D.A., and Kane, M.T. (1989). The effect of pre-letters on survey study response rates. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

D'Amico, R. (1990). The working world awaits: Employment experiences during and shortly after high school. In Wagner et al. (1990), Young people with disabilities: How are they doing? A comprehensive report from wavel of the National Longitudinal Transition Study of Special Education Students. Menlo Park, CA: SRI International.

DeStefano, L. and Metzer, D. (in preparation). Minimum competency testing and students in special education: An update on state level policy. Champaign, IL: Transition Institute, University of Illinois.

Dillman, D.A. (1978). Mail and telephone surveys. New York: Wiley-Interscience.

Donahoe, K. and Zigmond, N. (1990). Academic grades of ninth-grade urban learning-disabled students and low achieving peers. Exceptionality, 1, 17-28.

Edgar, E. (1987). Secondary programs in special education: Are many of them justifiable? Exceptional Children, 53(6), 555-561.

Edgar, E. (1988). Markers of effectiveness at the secondary level in special education. Proceedings of the research in education of the handicapped project director's meeting. Washington, DC: Office of Special Education Programs.

Fardig, D.B., Algozzine, R.F., Schwartz, S.E., Hensel, J.W., and Westling, D.L. (1985). Postsecondary vocational adjustment of rural, mildly handicapped students. Exceptional Children, 52, 111-121.

Flanagan, J.C. (1978). A research approach to improving our quality of life. American Psychologist, 33, 138-147.

Fowler, F.J. (1984). Survey research methods. Beverly Hills, CA: Sage Publications.

Frase, M. (1989). Dropout rates in the United States: 1988. Washington, DC: National Center for Education Statistics.

Freeman, R. and Medoff, J. (1982). Why does the rate of youth labor force activities differ across surveys? In Freeman, R. and Wise (eds.), The youth labor market problem: Its nature, causes, and consequences. Chicago IL: University of Chicago Press.

Halpern, A.S. (1987). A methodological review of follow-up and follow-along studies tracking school leavers from special education. Unpublished manuscript. Eugene, OR: University of Oregon.

Halpern, A.S., Nave, G., Close, D.W., and Nelson, D. (1986). An empirical analysis of the dimensions of community adjustment for adults with mental retardation in semi-independent living programs. Australia and New Zealand Journal of Developmental Disabilities, 12(3), 147-157.

Hammack, F.M. (1986). Large school systems' dropout reports: An analysis of definitions, procedures, and findings. In Natriello, G. (ed.), School dropouts: Patterns and policies. New York: Teachers College Press.

Hasazi, S.B., Gordon, L.R., and Roe, C.A. (1985). Factors associated with the employment status of handicapped youth exiting high school from 1978 to 1983. Exceptional Children, 51(6), 455-469.

Heal, L.W., and Chadsey-Rusch, J. (1985). The Lifestyle Satisfaction Scale (LSS): Assessing individuals' satisfaction with residence, community setting, and associated services. Applied Research in Mental Retardation, 6, 475-490.

Hoffman, K. (1980). Quality of life as perceived by persons who were classified as mentally retarded. Unpublished master's thesis. Lincoln, NE: University of Nebraska.

Javitz, H. and Wagner, M. (1990). National Longitudinal Transition Study of Special Education Students: Report on sample design and limitations, wave 1 (1987). Menlo Park, CA: SRI International.

Jay, E.D. (1990). A broader look at outcomes: Engagement in productive activities after high school. In Wagner, et al. (1990), Young people with disabilities: How are they doing? A comprehensive report from save 1 of the National Longitudinal Transition Study of Special Education Students. Menlo Park, CA: SRI International.

Keith, K.D. (1986). Quality of life in the community: Current status of adults with mental retardation. Paper presented at 110th annual meeting of the American Association on Mental Deficiency. Denver, CO.

Keith, K.D., Schalock, R.L., and Hoffman, K. (1986). Quality of life: Measurement and programmatic implications. Lincoln, NE: Region V Mental Retardation Services.

Kleinbaum, D.G., Kupper, L.L., and Miller, K.E. (1988). Applied regression analysis and other multi-variable methods. Boston: PWS-KENT Publishing Co.

Landesman, S. (1986). Quality of life and personal satisfaction: Definition and measurement issues. Mental Retardation, 24(3), 141-143.

Levin, E., Zigmond, N., and Birch, J. (1985). A follow-up study of 52 learning disabled students. Journal of Learning Disabilities, 18, 2-7.

Marder, C., and Cox, R. (1990). More than a label: Characteristics of youth with disabilities. In Wagner et al. (1990), Young people with disabilities: How are they doing? A comprehensive report from wave 1 of the National Longitudinal Transition Study of Special Education Studerts. Menlo Park, CA: SRI International.

Mithaug, D.E., Horiuchi, C.N., and Fanning, P.N. (1985). A report on the Colorado statewide follow-up survey of special education students. Exceptional Children, 51(5), 397-404.

Mithaug, D., Martin, J., Agran, M., and Rusch, F. (1988). Why special education graduates fail: How to teach them to succeed. Colorado Springs, CO: Ascent Publications.

National Council on Disability (1989). The education of students with disabilities: Where do we stand?: A report to the President and the Congress of the United States. Washington, DC: The Council, 1989.

Newman, L.A. (1990). Growing up, moving on: Aspects of personal and residential independence. In Wagner et al. (1990), Young people with disabilities: How are they doing? A comprehensive report from wave 1 of the National Longitudinal Transition Study of Special Education Students. Menlo Park, CA: SRI International.

Schalock, R.L., and Keith, K.D. (1986). Resource allocation approach for determining clients' need status. Mental Retardation, 24(1), 27-35.

Schalock, R.L., and Lilley, M.A. (1986). Placement from community-based mental retardation programs: How well do clients do after 8-10 years? American Journal of Mental Deficiency, 90(6), 669-676.

Schellenberg, S.J., Frye, D.W.M. and Tomsic, M.L. (1988). Loss of credit and its impact on high school students: A longitudinal study. Paper presented at the annual meeting of the American Educational Research Association. St. Paul, MN: St. Paul Public Schools.

Schroedel, J.G. (1984). Analyzing surveys on deaf adults: Implications for survey research on persons with disabilities. Social Science Medicine, 19(6), 619-627.

Semmel, D.S., Cosden, M.A., and Konopak, B. (1985). A comparative study of employment outcomes for special education students in a cooperative work placement program. Paper presented at the Council for Exceptional Children International Conference. Anaheim, CA.

Sitlington, P.L. (1986). The Iowa statewide follow-up. Unpublished manuscript. Iowa City, IA: University of Iowa College of Education.

Thorndike, E.L. (1939). Your city. New York: Harcourt, Brace and Co.

Thornton, H., Liu, M., Morrow, D., and Zigmond, N. (1987). Early identification of LD students at risk for becoming school dropouts. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

U. S. Department of Education (1989). "To assure the free appropriate public education of all handicapped children": Eleventh annual report to Congress on the implementation of the Education of the Handicapped Act. Washington, DC: U.S. Department of Education.

Valdes, K., Williamson, C. and Wagner, M. (1990). The National Longitudinal Transition Study of Special Education Students statistical almanac, volume 1: Overview. Menlo Park, CA: SRI International.

Wagner, M. (1990a). Reflections. In Wagner et al. (1990), Young people with disabilities: How are they doing? A comprehensive report from wave 1 of the National Longitudinal Transition Study of Special Education Students. Menlo Park, CA: SRI International.

b. Secondary school completion. In Wagner et al. (1990), Young people with disabilities: How are they doing? A comprehensive report from wave 1 of the National Longitudinal Transition Study of Special Education Students. Menlo Park, CA: SRI International.

c. Secondary school performance. In Wagner et al. (1990), Young people with disabilities: How are they doing? A comprehensive report from wave 1 of the National Longitudinal Transition Study of Special Education Students. Menlo Park, CA: SRI International.

Wagner, M. and Javitz, H. (1990). National Longitudinal Transition Study of Special Education Students: Measurement and analysis issues. Menlo Park, CA: SRI International.

Wagner, M., Newman, L. and Shaver, D. (1989). The National Longitudinal Transition Study of Special Education Students: Report on procedures for the first wave of data collection (1987). Menlo Park, CA: SRI International.

Wagner, M.M. and Shaver, D.M. (1989). Educational programs and achievements of secondary special education students: Findings from the National Transition Study. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Wehman, P., Kregel, J., and Seyfarth, J. (1985). Employment outlook for young adults with mental retardation. Rehabilitation Counseling Bulletin, 90-99.

Williams, P., and MacDonald, A. (1986). The effect of non-response bias on the results of two-stage screening surveys of psychiatric disorder. Social Psychiatry, 21, 182-186.

Worthen, B.R., and Sanders, J.R. (1987). Educational evaluation: Alternative approaches and practical guidelines. New York: Longman.

Zautra, A., and Goodhart, D. (1979). Quality of life indicators: A review of the literature. Community Mental Health Review, 4(1), 1-10.

Zigmond, N., and Thornton, H. (1985). Follow-up of postsecondary age learning disabled graduates and dropouts. Learning Disabilities Research, 1(1), 50-55.

Zigmond, N. and Thornton. H. (1985). Learning disabled graduates and dropouts. Learning Disabilities Research, 1(1), 50-55.